NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

TYPHOON MOTION FORECASTING USING EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS OF THE SYNOPTIC FORCING

by

Alan R. Shaffer

March 1982

Thesis Advisor: R. L. Elsberry

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Typhoon Motion Forecasting Using Empirical Orthogonal Function Analysis of the Synoptic Forcing		5. TYPE OF REPORT & PERIOD COVERED Master's thesis; March 1982
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Alan R. Shaffer		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1982
		13. NUMBER OF PAGES 150
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Typhoon Motion Forecasting

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Empirical Orthogonal Function (EOF) analysis is used to describe the synoptic forcing features of selected northwestern Pacific Ocean tropical cyclones from 1967 to 1976. EOF analysis is applied to the geopotential field at 850, 700 and 500mb on a 120 point grid with 5 degree latitude and longitude spacing that is centered on the storm. The 120 EOF coefficients (for each level) are computed for a sample of 454 cases in the history file.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

#20 - ABSTRACT - (CONTINUED)

The coefficient vectors are truncated to the first 10 coefficients, based on the Monte Carlo selection criteria of Preisendorfer and Barnett. These coefficients describe about 85% of the variance in the fields. The synoptic forcing represented by the EOF coefficients is then used as a predictor in a regression analysis track forecast scheme, along with past storm movement and intensity during the past 36 hours. The EOF-based regression equations are verified over an independent sample of 50 storms, and the position errors compared to the official Joint Typhoon Warning Center (JTWC) forecast errors. The EOF-based regression equations give, on the average, a 17% reduction in error when compared to the official forecast issued by JTWC. Over the independent sample, the 500mb equations performed better than the equations of the other two levels.

Typhoon Motion Forecasting Using Empirical Orthogonal Function Analysis of the Synoptic Forcing

by

Alan R. Shaffer
Captain, United States Air Force
B.S., University of Vermont, 1976
B.S., University of Utah, 1977

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
March 1982

ABSTRACT

Empirical Orthogonal Function (EOF) analysis is used to describe the synoptic forcing features of selected northwestern Pacific Ocean tropical cyclones from 1967 to 1976. EOF analysis is applied to the geopotential field at 850, 700 and 500mb on a 120 point grid with 5 degree latitude and longitude spacing that is centered on the storm. The 120 EOF coefficients (for each level) are computed for a sample of 454 cases in the history file. The coefficient vectors are truncated to the first 10 coefficients, based on the Monte Carlo selection criteria of Preisendorfer and Barnett. These coefficients describe about 85% of the variance in the fields. The synoptic forcing represented by the EOF coefficients is then used as a predictor in a regression analysis track forecast scheme, along with past storm movement and intensity during the past 36 hours. The EOF-based regression equations are verified over an independent sample of 50 storms, and the position errors compared to the official Joint Typhoon Warning Center (JTWC) forecast errors. The EOF-based regression equations give, on the average, a 17% reduction in error when compared to the official forecast issued by JTWC. Over the independent sample, the 500mb equations performed better than the equations of the other two levels.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I wish to thank Professor Russ Elsberry for patience and guidance through the course of the research. He has the greatest ability a teacher may have; that of guiding a student to reach for higher potential and learning. Additionally, I wish to thank Professor R. L. Haney for the valuable comments on the manuscript. The comments were instructive and very greatly appreciated.

I would also like to thank Dr. Rudolph Priesendorfer, for taking time out of a very busy schedule to discuss aspects of the research during the formative stages.

A hearty thanks to Major Dan Brown, USAF, who accomplished the initial research using this data, and in so doing encountered and solved the data extraction problems. Additionally, many of the computer programs used during this research were simply modifications of programs written by Dan. His initial work greatly reduced the number of problems.

Additionally, the staff of the W. R. Church computer center and in particular Kristina Butler and Jeaneanne Washington for all their aid in using the computer. Their aid made coping with the system much easier, and made time spent with the system much more profitable.

To fellow students Robert Allen, Charles Hopkins and Barry Donovan and the rest for discussing and offering critiques of the research as time progressed; thank you one and all.

Finally, I would like to dedicate this work to the memory of my father, Leonard Shaffer, through whom all things became possible.

I. INTRODUCTION

Tropical storms spawned over the western North Pacific Ocean genesis region have great impact on both civilian and military populations; accurate movement forecasts are critical to reduce their impact upon these communities. The Joint Typhoon Warning Center (JTWC), Guam, Marianas Islands, issues the official forecast (to United States military agencies) of tropical storm movement and intensity for storms generated in this region. Using current forecast techniques, these official forecasts have an average forecast position error on the order of 120, 240 and 360 nautical miles for 24-, 48-, and 72-hour forecasts (Annual Typhoon Report, JTWC, 1981). There is potential for improvement.

Present forecast techniques for tropical storm movement may be generally categorized as being either statistical (which includes analog techniques) or dynamical. The motivation driving the two types of forecasts differs greatly. Statistical forecasts typically use regression or analog methods with all available historical storms having archived data to produce a statistically optimal position forecast. Regression analysis methods assume that certain variables deterministically correlate with future storm displacement. These correlated variables are then used in a regression analysis to produce a forecast. Leftwich and Neumann (1977), for example, use a second order polynomial regression with seven primary predictors

to forecast typhoon movement. The seven predictors include Julian date, initial latitude and longitude, and past 12- and 24-hour zonal and meridional movement. Since they used polynomial regression, these seven primary predictors actually give rise to 35 predictors when the second order predictors are formed. Using these predictors, Leftwich and Neumann were able to account for 65% of the variation in the zonal displacement and 53% of the variation in the meridional displacement for 12 hours. Over a 72-hour period, the amount of explained variance became progressively smaller. Analog techniques (e.g., Jarrell and Sommervell, 1970), use the historical file of storms to identify storms, and the surrounding environmental fields, that have strong similarities to the present storm. Then, a weighted similarity index of certain variables is used to select those storms in the history file that are most similar to the present storm. A weighted average of the selected storm tracks is the basis of the forecast movement of the present storm. The justification for using this technique is that a storm with similar location and surrounding fields should also have a similar track. Jarrell and Sommervell (1970) present an analog scheme which is the original version of the scheme used presently at JTWC.

In contrast to the statistical methods, dynamic forecast techniques assume that the motion of the storm may be forecast directly from numerical integration of geophysical governing equations (momentum, continuity and thermodynamic

equations, for example). Harrison (1973) presents a simple nested grid model to forecast typhoon movement using the primitive equations. This is the original version of the operational nested tropical cyclone model available at JTWC (Harrison, 1981).

Both statistical and dynamical forecast methods have weaknesses. The statistical methods have two primary problems; first, since they are based on historical data cases, any storm that has an unusual motion is not likely to be forecast well. Additionally, the use of statistical methods tends to homogenize (smooth) the forecast. Forecasts using a blend of similar past history storms are typically insensitive to subtle differences in the synoptic (dynamic) forcing fields. Thus, purely statistical methods have deficiencies in forecasting the unusual case and inability to distinguish subtle differences in the synoptic-scale fields.

Dynamic forecasts, on the other hands, have limitations in both theory and cost. Due to the smallness of the coriolis parameter in tropical regions, a geostrophic relationship is not feasible. This makes initialization of fields difficult and increases the probability that any erroneous data points will deteriorate the numerical forecast rapidly. Convective heating is a driving mechanism for development of tropical storms, rather than baroclinic instability as in the mid-latitudes. Unfortunately, convective heating is very difficult to model (Haltiner and Williams, 1980). Therefore, the governing equations are suspect in the tropics, due to poor

initialization and modeling of convective heating. An even greater problem is that interaction between different scales of motion is critical to maintain an energy balance in the tropical cyclone. If the grid spacing is not small enough, the energy balance will be altered, and possibly give spurious solutions. For this reason, the grid must have very fine resolution to simulate numerically this interaction. The cost of numerical integration on a fine grid can be very large due to the Courant-Fredrichs-Levy (CFL) condition which requires smaller integration time steps as the grid spacing decreases (Haltiner and Williams, 1980). An additional problem with a fine grid model is that there are generally inadequate wind and mass observations to initialize the numerical model in the tropics, and this problem is increased as the grid size is reduced.

With the difficulties in both types of forecasting methods, an alternative method is proposed here. This study will employ Empirical Orthogonal Functions (EOF's) to represnet numerically the large scale synoptic (dynamic) fields. Then, these functions will be used to forecast the tropical storm movement using regression equations. This approach is novel for forecasting of tropical storm movement, in the sense that previous regression analysis methods (Leftwich and Neumann, 1977, for example) have not incorporated the entire synoptic forcing field. If an attempt to develop a simple linear regression model using a large synoptic field is made, the number

of predictors becomes prohibitive, as each grid point value relative to the storm would be a predictor. Early analog studies used only a single feature from the synoptic chart, such as the 700mb trough longitude to the north of the storm, to represent the synoptic field. This study will use the Empirical Orthogonal Function representation of the entire synoptic forcing field around the tropical storm. Therefore, in a broad sense, this approach may be thought of as a dynamically-based statistical forecast scheme. This type of approach is not totally without precedence. Lorenz (1977) states:

> In an informal conversation in which this writer (Lorenz) took part in about 20 years ago, the question arose as to how the best system for producing the operational objective 24 h prog could be developed, if the system had to be ready within one year. We more or less agreed that the further improvements in numerical weather prediction to be expected in a single year would be small, and that the greatest gains would come from an empirical scheme in which the numerically produced prognostic charts, or "numerical progs" would enter as predictors....

Substitution of "improved tropical forecast scheme" for "24 h prog" in the quotation gives the basis and purpose of this study.

Empirical Orthogonal Function analysis allows a field with many grid points to be represented by a linear combination of a few constant vectors and variable coefficients, while retaining a large portion of the total variation (from the mean state) in the field. Thus, a synoptic field with many grid points may be accurately represented by only a few variable

coefficients (given the vectors are constant), which makes the technique ideal to use with regression analysis. For example, Kutzbach (1967) was able to represent 88% of the total variation in average January temperatures at 23 stations (grid points) in North America over a 25-year period by using only five coefficients and constant vectors. That is, the entire synoptic scale chart of mean temperature was represented by a 23 element vector, and all of the data were stored in 25 individual 23-element vectors. Thus, Kutzback was able to reduce the number of vectors needed to describe the January temperature field for each year (at the 23 locations) from 25 to 5.

The Empirical Orthogonal Function analysis in this study is used for data reduction and representing synoptic fields numerically. The synoptic-scale forcing upon the tropical storm may be represented by only a few coefficients obtained from the analysis. These coefficients may be then used to forecast statistically the tropical storm movement. In this manner, the synoptic (dynamic) forcing is incorporated into the statistical forecasting scheme. Thus, the primary purpose of this study is to investigate the role of the synoptic forcing and to forecast tropical storm movement from this forcing.

II. DATA ACQUISITION AND FIELD DEFINITION

The tropical cyclone tracks and height data used in this study are identical to those used by Brown (1981). The data required for an individual case include D-value fields at 850, 700 and 500mb and the storm location history prior to and after the forecast time. A relocatable 120-point grid is defined with 5-degree grid spacing in both longitude and latitude. The grid covers an areal extent of 70 degrees east to west and 35 degrees north to south. Individual grid points are numbered as shown in Fig. 2-1. The grid is moved each map time such that the tropical storm is always located at grid point 70. A moveable grid can create difficulty in obtaining composite variable fields due to the longitude convergence as the storm moves further north. For this study, this problem is assumed to be of minor importance, and any composite type fields are computed assuming a flat earth. It will be shown below that this assumption is not too bad over the domain used in this study.

D-values are defined (Huschke, 1959) as height deviations (in meters) from the standard atmosphere height at a constant pressure surface, and are typically positive in the tropics. The source of the data is the operational Fleet Numerical Oceanography Center's (FNOC) Northern Hemisphere (63 X 63) analyses at 850, 700 and 500mb. The following selection conditions are required:

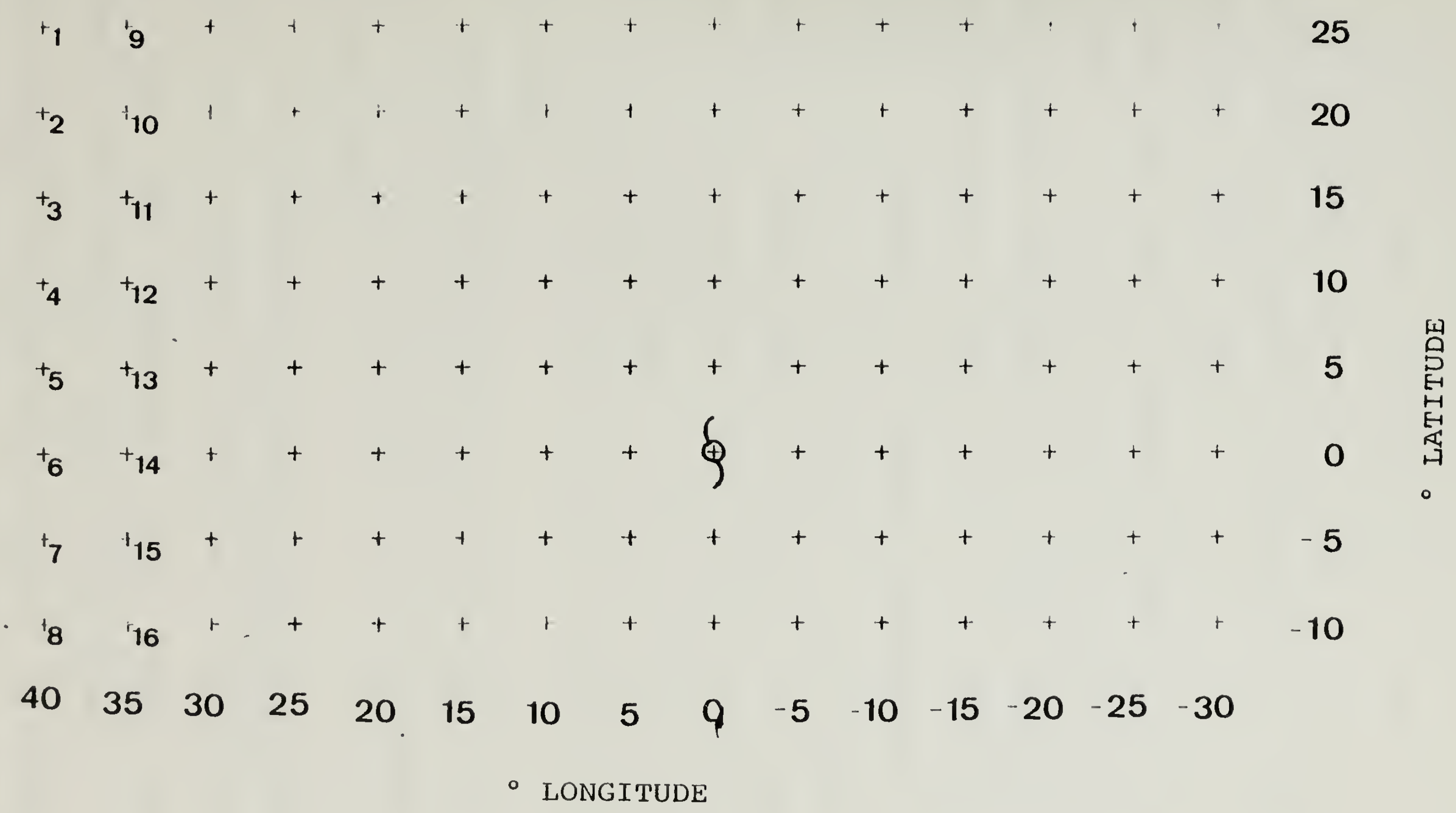


Fig. 2-1. The moveable 120 point grid on which the D-values were extracted relative to the position of the storm. The storm is located at grid point 70, denoted by § . Distances in degrees latitude and longitude to the various grid points are shown. The grid point numbering system is demonstrated in the first two columns.

(1) A tropical cyclone of at least tropical storm (35 knots) intensity must be present west of 180°W;

(2) The storm must persist at least 30 hours with tropical storm intensity or greater, as analyzed by the Joint Typhoon Warning Center (JTWC), Guam;

(3) The storm must be located between 10° and 25°N. This requirement was included to insure the grid did not extend into the Southern Hemisphere, and was not comprised of primarily mid-latitude D-values. Since the latitudinal domain is limited, the problem of longitude convergence is not a significant problem at the latitudes of the domain. The distance from the western edge of the grid to the storm ranges from 1772 nautical miles at 10°N to 1631 nautical miles at 25°N, to 1474 nautical miles at 35°N and finally to 1157 nautical miles at 50°N. This range of distance is considered insignficicant.

(4) Since the storm position is coupled with the upper level analysis, only storms existing at 0000 GMT and 1200 GMT are considered;

(5) A 36-hour separation between subsequent positions of the same storm is required to provide a pseudo-independence between cases. This independence is a critical consideration whenever statistical analysis is conducted.

After defining the selection criteria (1) through (5), the JTWC Annual Typhoon reports from 1967 to 1976 were examined to select potential cases. These particular years were chosen

because the FNOC Northern Hemispheric D-value fields were available from Systems and Applied Sciences, Monterey, California, during these years. Examination of the JTWC reports yielded 560 potential cases meeting the criteria above. However, only 540 cases had the required D-value data. Of these 540, there were data problems with an additional 36 cases, leaving 504 valid cases. Archived D-value data were interpolated to the 120-point movable grid by the method of Bessel linear interpolation (Brown, 1981). The phrase "base time" will be used to define the time of the initial D-value field, and therefore the forecast. The storm warning position from JTWC is used as the location at the base time and at all times prior to the base time, whereas the JTWC best-track position is used for verification positions. This is a significant difference from Brown (1981), who used only the best-track positions for all historical locations. Warning positions are used because they are the actual locations available at the time of forecast. The best-track positions are calculated after the typhoon season, and are not available to the forecaster in the field. Nevertheless, they are assumed to be the optimal position and therefore the value that the forecast scheme tries to replicate.

Storm warning positions are obtained at the base time and 12, 24 and 36 hours prior to the base time. Best track positions are gathered for future positions in 6-hour increments from the base time to 84 hours in the future. Therefore, a storm with complete history has continuously available locations

for 120 consecutive hours. The set of three levels of D-value fields, four warning positions and 15 best-track positions comprise the entire set for each case. The number of cases having X available prior warning positions and Y future best track locations available is shown in Table 2-1. It is interesting to note that while there are 504 valid cases meeting criteria (1) through (5), only 401 cases have all 36-hours of prior warning position. Furthermore, only 185 cases have both 36-hours prior warning position and 84 hour future best track positions available. The number of storms with 36-hour prior warning position available increases to 298 available cases with 48-hour future best track location and 401 storms with 30-hour future best track locations at tropical storm strength. The number of cases with a full 36-hour history is important when the regression equations are developed.

The composite D-value fields at 500, 700 and 850mb using all 504 cases are shown in Figs. 2-2, 2-4 and 2-6. Of interest is the relatively small gradient in the tropics in the 500mb composite. This level has relatively little indication of a tropical disturbance at grid point 70, since the 500mb level is near the level at which the surface cyclone becomes an upper-level anticyclone. The lower level (850 and 700mb) charts show fairly strong gradients in the D-value field around point 70. Figs. 2-3, 2-5 and 2-7 show the D-value standard deviations for all three levels. As expected, the greatest D-value variation is near the storm location and in the mid-latitude westerlies to the north. These mean and standard

TABLE 2 - 1

The number of valid cases by prior JTWC warning positions and future JTWC best track position. See text for details.

FUTURE LOCATIONS AVAILABLE (in hours)	TOTAL	NUMBER OF CASES WITH BASE TIME AND 12 HOUR PRIOR WARNING POSITIONS ONLY	24 HOUR	36 HOUR
6	504	461	422	401
12	504	461	422	401
18	504	461	422	401
24	504	461	422	401
30	504	461	422	401
36	480	439	400	379
42	380	351	315	298
48	380	351	315	298
54	380	351	315	298
60	352	325	291	274
66	265	242	215	200
72	265	242	215	200
78	265	242	215	200
84	265	221	199	185

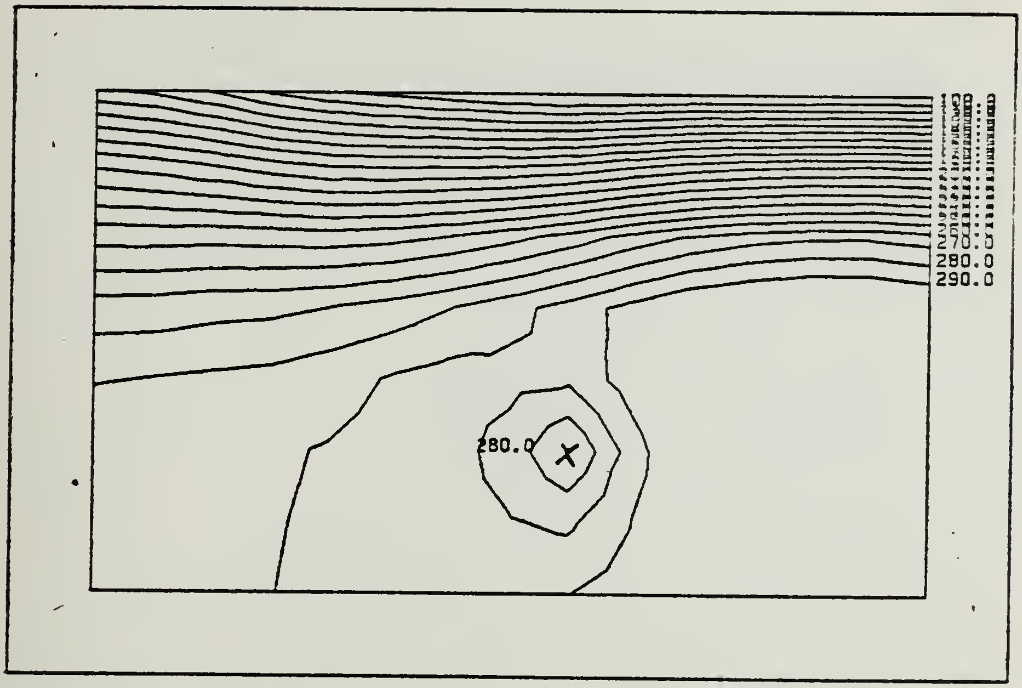


Fig. 2-2. The mean (composite) D-value field at 500mb. Isopleths are deviation in meters from standard atmosphere. Storm is always located at grid point 70 (X).

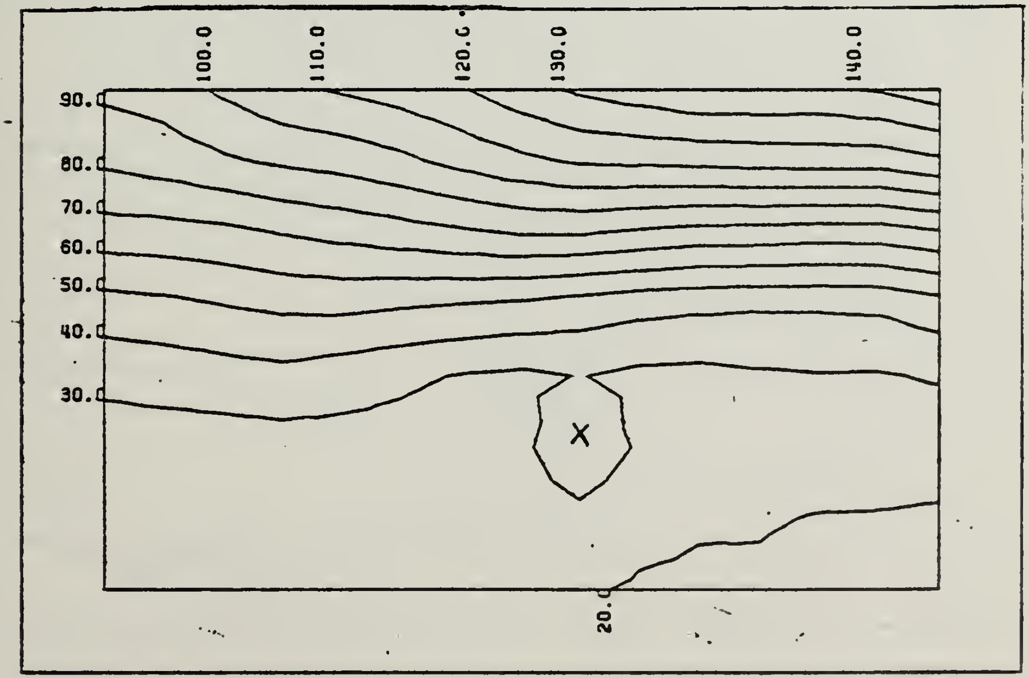


Fig. 2-3. The composite standard deviation D-value field (in meters) at 500mb. The storm is always located at grid point 70 (X).

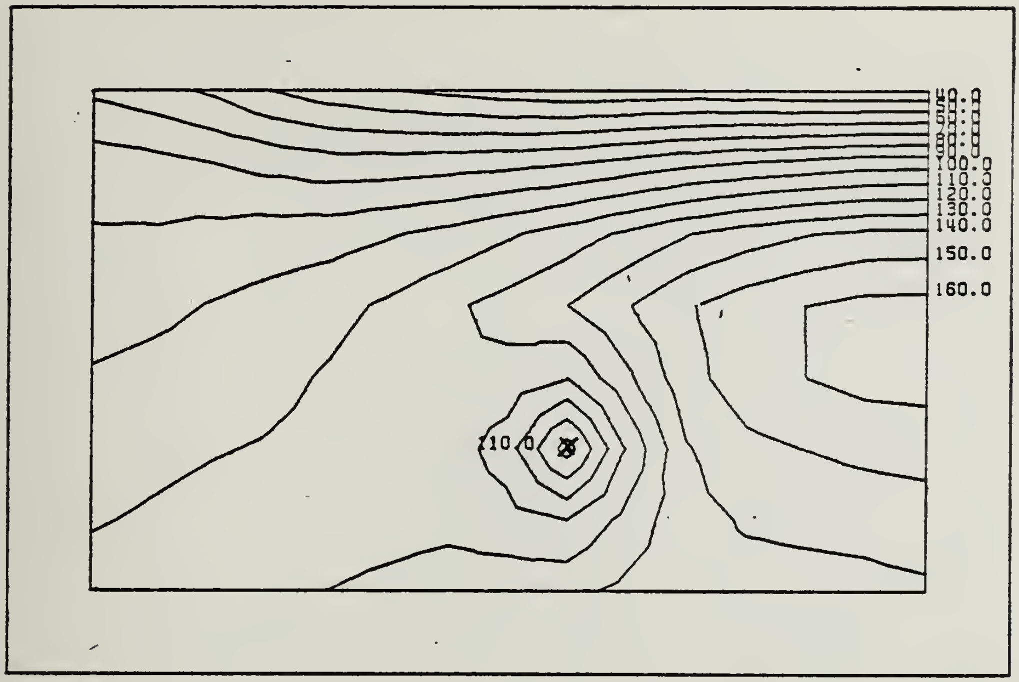


Fig 2-4. Similar to Fig. 2-2, except for 700mb.

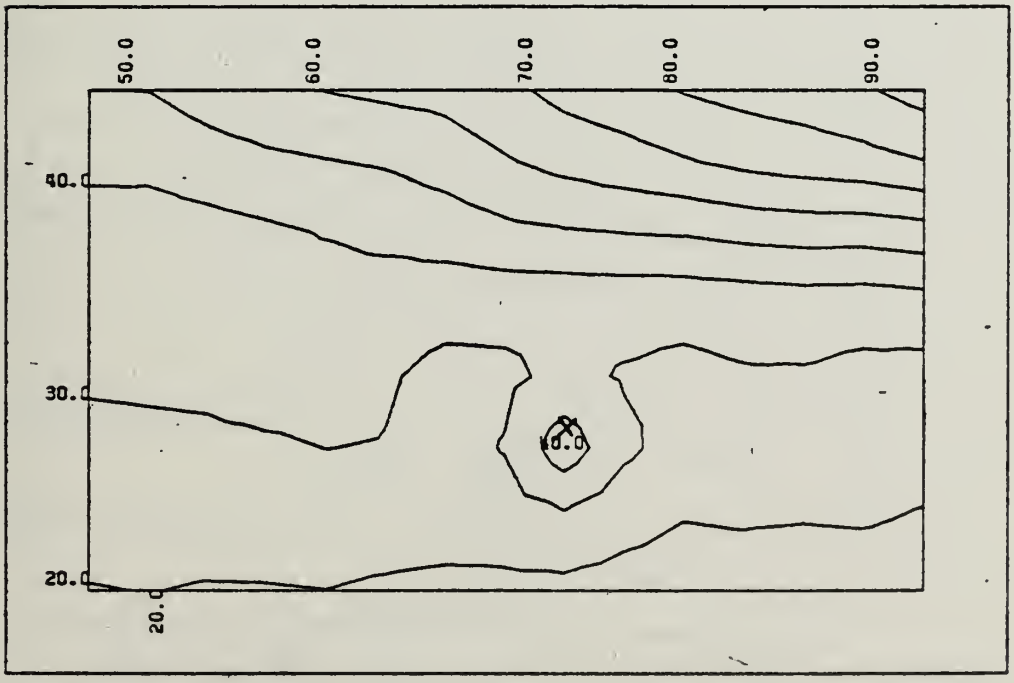


Fig. 2-5. Similar to Fig. 2-3, exept for 700mb.

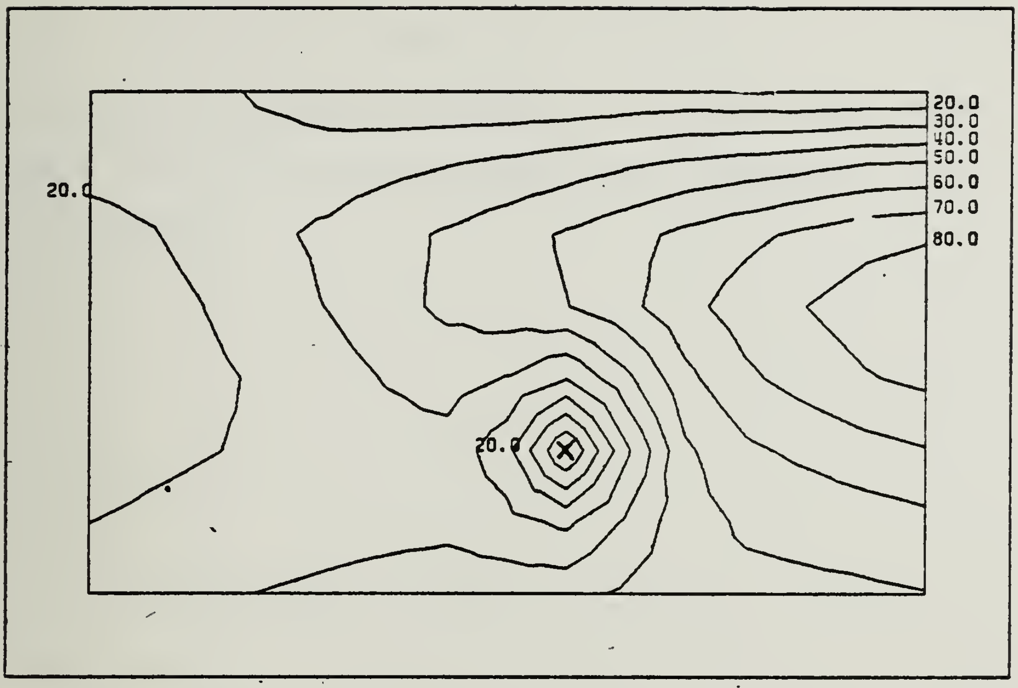


Fig 2-6. Similar to Fig. 2-2, except for 850mb.

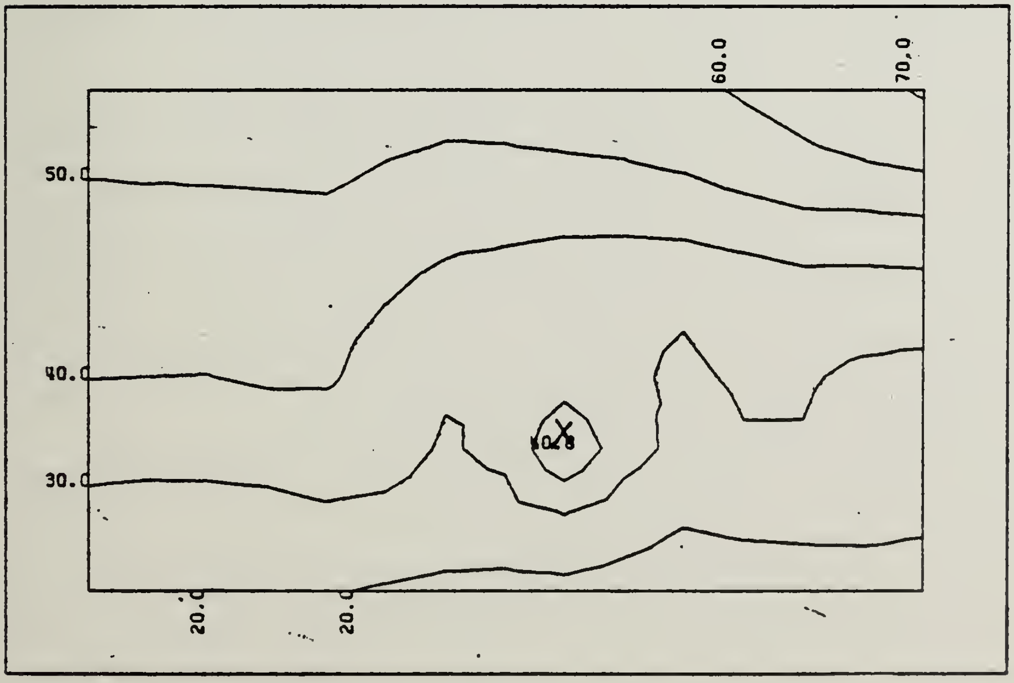


Fig. 2-7. Similar to Fig. 2-3, except for 850mb.

deviation fields are the fields used in normalizing the data for each case, by grid point, for use in the Empirical Orthogonal Function analysis. The 504 cases comprise the data set from which the Empirical Orthogonal Functions will be obtained.

III. EMPIRICAL ORTHOGONAL FUNCTIONS

A. BACKGROUND

The terminology "Empirical Orthogonal Function" (EOF) was introduced by Lorenz (1956). Actually, EOF analysis is a variation of the statistical technique of principal components, and was introduced in its current form by Hotelling (1933), and was based on an idea of Pearson (1901). Before delving into the mechanics of EOF analysis, the basic concepts and meaning of principal components will be presented geometrically. Geometric meanings presented for principal components are valid for EOF's, since EOF's differ from principal components only by a scaling factor.

Principal components aid in explaining interrelations of individual variables acting on a larger field. Morrison (1967) presents a concise geometric interpretation of the method. Principal components may be drawn from data sets in any number of dimensions, but their meaning is most easily seen in three-dimensional space. Suppose three variables (X_1,X_2,X_3) form a trivariate observation space. For example, X_1, X_2, and X_3 could be the 500mb D-value at gridpoints 1, 2 and 3 respectively. A large collection of simultaneously measured values of the three variables could be plotted as in Fig. 3-1. The shaded ellipsoid in the figure represents the scatter plot of many observations of the three variables. The origin of the axis is the mean value for each of the three variables. The

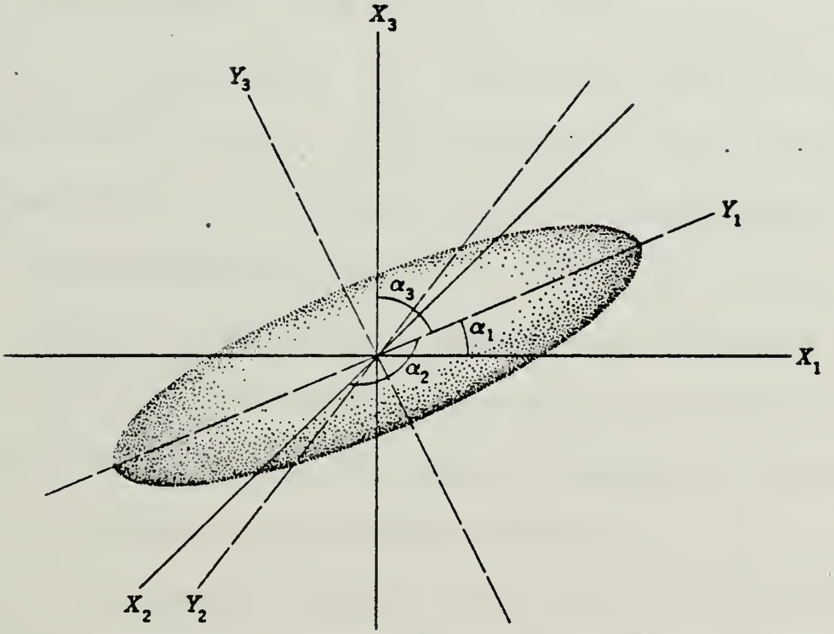


Fig. 3-1. An example of trivariate principal components. See text for details (Morrison, 1967).

first of the three principal components (there will generally be three unique principal components in three dimensions) is the major axis of the ellipsoid, denoted as Y_1 in the figure. In other words, the first principal component is the axis in space that explains the maximum variation from the origin in the three-dimensional space. For this reason, the term principal axes is sometimes used instead of principal components. It is easily seen that this first principal component can be represented by a vector (and the vector 180 degrees out of phase) originating at the origin. The second principal component is the minor axis (Y_2) which describes the maximum amount of variation in the ellipsoid that is not explained by the first component. The second principal component is also subject to the constraint that it be orthogonal to the first component. This is identical to saying the second principal component is the largest minor axis which is orthogonal to the major axis. The third principal component is the third minor axis (Y_3) which explains the remainder of the variation of the ellipsoid. This component is subject to the constraint that it be orthogonal to the first two components (axes). Thus the three principal components explain the total variation in the observation ellipsoid. The components are simply orthogonal axes, in three dimensions. It is seen from this simplified example that the technique may be easily extended to application in multiple dimensions. If the axes are defined by vectors, it is straightforward to find orthogonal vectors by

standard methods. This orthogonality constraint simplifies identification and interpretation.

In M-dimension space, there will be M (or occasionally fewer) orthogonal components, which are simply the orthogonal vectors in M space. If there are fewer than M unique components, the observation variables are overdefined, and two or more of the describing variables are perfectly correlated. If this is the case, one of these perfectly correlated variables may be omitted with no loss of information.

As mentioned, Lorenz (1956) introduced the terminology "Empirical Orthogonal Function", and made the application to the atmospheric sciences. The mathematical method used for finding the orthogonal components or vectors involves solution of the eigenvalue problem in M space. EOF's are simply principal components that have not been scaled by the square root of the corresponding eigenvalue found in the solution. This subtle difference is really of little concern. It does cause a slight modification in the computations, and also slightly changes the interpretation of the results. This interpretation difference arises because the loadings (elements) of the solution eigenvector (principal component) are nothing more than the correlation of the variables in a given dimension with the principal axis it defines (Anderson, 1958). No such easy interpretation of the loadings is possible with EOF's. This modification is not significant, and the salient points and geometric interpretation valid for principal components are likewise valid in EOF analysis; only the lengths of the

orthogonal vectors are different. The mathematical details will be covered in the next section.

EOF analysis normally has been used in two primary applications in geophysical sciences. These are either as a map-typing tool, or as a tool for reducing dimensionality and explaining the variance structure of a large field. For example, Stidd (1967) uses EOF analysis to describe the variation in average monthly rainfall in Nevada. In this paper, Stidd states:

> eigenvectors might be regarded as an ultimate development in the use of orthogonal functions to describe patterns or arrays of data.

He goes on to show that annual precipitaion in Nevada may be described primarily by one of three basic "components". The three are: (1) a winter maximum from large scale storms; (2) a secondary peak during the summer due to thunderstorms; and (3) a small effect due to the removal and inclusion of water into the hydrological structure due to snow pack. EOF analysis allows extraction of each component and allows the researcher to determine the primary variables driving each of the components. Additionally, by using a linear combination of the eigenvectors (components), it is possible to determine and estimate the rainfall amount in data sparse and non-observed regions. This estimation is done by interpolation of coefficients associated with each eigenvector. These coefficients will be explained more fully in the next section. Stidd was able to explain 93% of the total variance in the annual rainfall in Nevada by using only three eigenvectors and coefficients.

This is compared to the initial estimation which required 12 charts (one for each month). The key points are that Stidd was able to both isolate the causes behind annual variation in Nevada rainfall (over all locations in Nevada), and additionally, reduce the data required to make this estimate by 75% (from 12 charts to three). This "gleaning of the forcing pattern" and data reduction use of EOF's has been used frequently in meteorological applications. Other examples of EOF use in this manner are found in Rinne and Karhila (1979), and Craddock and Flood (1969).

Another application of EOF analysis has been for map typing. Brown (1981) uses EOF analysis to divide a large sample of cases into smaller discrete subsets by map typing based on the coefficients derived from EOF analysis. The primary objective was to use the subsets of similar cases to form analogue-type forecasts of tropical cyclone tracks. Accuracy of forecasts using this map typing scheme is generally less than with other objective tropical cyclone motion forecasting techniques.

B. MECHANICS OF THE EOF METHOD

The mechanics of EOF analysis presented here follows an elegant treatment by Kutzbach (1967). The notation used in this development is defined as follows; a single underscored variable in lower case letters is a vector (e.g., $\underline{e}$), an uppercase variable with two underscores is a matrix ($\underline{\underline{A}}$), and a primed vector of matrix is the transpose ($\underline{e}'$). The raw data field (in this study, the 120 grid point fields of

D-values) is formed into a matrix, $\underline{\underline{A}}$. This matrix is constructed so that each column consists of the 120 observed D-values for a particular data case. Each row represents the D-values at the same grid point for all data cases. If there are N separate data cases (storms), with each case having M grid point values, $\underline{\underline{A}}$ is an M X N matrix representing the observed D-value fields. The objective of EOF analysis is to determine the single vector ($\underline{e}$) in M dimensions that best represents all of the N observation vectors. This is equivalent to saying that one wants to find the vector ($\underline{e}$) that minimizes the summed squared error of all observation vectors compared to ($\underline{e}$). Therefore, EOF analysis may be thought of broadly as a multi-dimensional extension of a least squares technique.

The matrix $\underline{\underline{A}}$ may be constructed in one of three ways: with the actual data values; with the departure from mean data values; or with the normalized departure from mean values. There are advantages and disadvantages to using each type of initialization for the data matrix $\underline{\underline{A}}$. In the first case, the resultant EOF's will have magnitudes on the order of the actual data, and will effectively represent the actual component field. Morrison (1967) points out that this type of input matrix may be dangerous to use if the variables in the different dimensions vary widely in magnitude. As seen in the mean and standard deviation charts of the fields (Figs. 2-2 through 2-7), this could be a problem here, since the D-values are generally quite a bit lower in the northern portion of the grid,

as well as having larger variation in the north. There are systematic differences in magnitude at different points on the grid (dimensions). Thus, the grid points with larger values are given more weight than the grid points with smaller values, and some of the meaning of the resultant eigenvectors is lost. For this reason, this type of input data was not used. A second potential form for the data matrix $\underline{\underline{A}}$ is to have the elements be comprised of the deviations from the mean value of a given dimension (row). This type of approach is more in line with the classical principal components approach. In this case, the eigenvectors are extracted from the covariance matrix. This is really the main advantage to this form, while the primary disadvantages are that the interpretation of the resultant eigenvectors becomes muddled due to scaling of the dimensions and again, there is not equal weight between dimensions if their magnitudes differ. The third choice for the input data matrix form is to use normalized departures from the mean. This has a disadvantage in that it may smooth slightly the resultant eigenvectors (Kutzbach, 1967). This approach was selected because the variations in all dimensions are equally weighted in extracting the eigenvectors. In this study, normalization is accomplished by subtracting the mean value at that grid point (over all cases), and then dividing by the standard deviation of that grid point over all cases;

$$(a_{mn})_T = (a_{mn} - \bar{a}_m)/s_{am}$$

where:

$(a_{mn})_T$	is the transformed data point
a_{mn}	is the original data point (D-value)
$\bar{a}_m$	is the mean of a at grid point m (taken over all n cases)
s_{am}	is the standard deviation of a at grid point m (over all n cases).

Brown (1981) discusses in more detail various methods of normalization transformations.

After obtaining the normalized input data matrix $\underline{\underline{A}}$ (over all N cases), the next step is to maximize the quantity

$$(\underline{e}'\underline{\underline{A}})^2 N^{-1} / \underline{e}'\underline{e} \ , \qquad (1)$$

(where, unless otherwise noted, any product of two vectors or matrices is the dot (inner) product) under the constraint that

$$\underline{e}'\underline{e} \ = \ 1. \qquad (2)$$

Equation (1) is the squared product of an arbitrary vector ($\underline{e}$) and the actual data vectors. Constraint (2) is made simply to normlaize the maximized product. This maximization of (1) with constraint (2) may be rewritten:

$$\text{Max}\{y: \underline{e}'e = 1\} \quad \text{where} \quad y \ = \ (\underline{e}'\underline{\underline{A}})^2 \ N^{-1}, \qquad (3)$$

or

$$\text{Max}\{y: \underline{e}'\underline{e} = 1\} \quad \text{where} \quad y = \underline{e}'\underline{\underline{A}}\,\underline{\underline{A}}'\,\underline{e}N^{-1}, \qquad (4)$$

Defining $\underline{\underline{R}} = \underline{\underline{A}}\,\underline{\underline{A}}'\,N^{-1}$, equation (4) may be written as

$$\text{Max}\{y: \underline{e}'\underline{e} = 1\} \quad \text{where} \quad y = \underline{e}'\,\underline{\underline{R}}\,\underline{e}\ . \qquad (5)$$

It is of interest to note that the form of $\underline{\underline{R}}$ is the cross product matrix if $\underline{\underline{A}}$ is comprised of the actual data. However, $\underline{\underline{R}}$ is the covariance matrix, or the correlation matrix, if the input matrix $\underline{\underline{A}}$ has elements which are deviations from the mean or normalized deviations from the mean, respectively. Premultiplying both sides of equation (5) by $\underline{e}$ results in

$$\underline{e}\,y = \underline{\underline{R}}\,\underline{e}\ . \qquad (6)$$

Morrison (1967) shows that maximization of y leads to the requirement that $|\underline{\underline{R}} - y\underline{\underline{I}}| = 0$, or else the solution is trivial. Maximization of (6), therefore, yields the eigenvalue problem, where y is the eigenvalue.

Equation (6) applies to maximization of one eigenvector only. Since there are M dimensions in the original problem, one wishes to maximize the explained variance in each of the dimensions. Therefore, it is convenient to rewrite (6) for all vectors in the M-space as

$$\underline{\underline{E}}\,\underline{\underline{Y}} = \underline{\underline{R}}\,\underline{\underline{E}}\ . \qquad (7)$$

Here, $\underline{\underline{E}}$ is an M X M matrix, rather than a vector as was the case for (6). It turns out that the elements of $\underline{\underline{Y}}$ are the eigenvalues found solving $|\underline{\underline{R}} - \underline{\underline{Y}}\,\underline{\underline{I}}| = 0$. Each column of $\underline{\underline{E}}$ is an eigenvector associated with a single eigenvalue Y_i. It follows from the definition of eigenvectors that they are orthogonal (uncorrelated). Again, the necessary condition in finding $\underline{\underline{E}}$ is that $\underline{\underline{E}}'\underline{\underline{E}} = \underline{\underline{I}}$, the identity matrix.

Returning to the basic definition of $\underline{\underline{R}}$, it is seen by substitution that

$$\underline{\underline{E}}'\underline{\underline{A}}\,\underline{\underline{A}}'\underline{\underline{E}} = N\,\underline{\underline{Y}}\ . \qquad (8)$$

Morrison (1967) has shown that the eigenvector associated with the largest eigenvalue (y_1) is the vector that explains the maximum variation in $\underline{\underline{R}}$. In fact, the first eigenvector explains

$$y_1 / \sum_{i=1}^{m} y_i \qquad (9)$$

of the total variation in $\underline{\underline{R}}$. The variance unexplained by the first (largest) eigenvector is the residual. The second eigenvector is associated with the second largest eigenvalue, and explains the maximum variation remaining in the residual field, and is given by

$$y_2/\sum_{i=1}^{m} y_i \;. \tag{10}$$

Therefore, the first two eigenvectors together explain

$$y_1 + y_2/\sum_{i=1}^{m} y_i \tag{11}$$

of the total variation in $\underline{\underline{R}}$. The process continues with each successive eigenvector describing the maximum remaining variation in the residual field. The final eigenvector is simply any variation in the total mean field left unexplained by the combination of all previous eigenvectors. As the last eigenvector explains all of the remaining variation in the field, the total variation in $\underline{\underline{R}}$ is explained by all of the eigenvectors.

Any of the original fields (cases) may be obtained by calculating the EOF coefficients. These coefficients (called multipliers by Stidd, 1967, and others) are also orthogonal and are found by defining:

$$\underline{\underline{C}} = \underline{\underline{E}}'\underline{\underline{A}} \;, \tag{12}$$

where $\underline{\underline{C}}$ is an M X N matrix. The nth row of the coefficient matrix ($\underline{\underline{C}}$) is the orthogonal coefficient vector corresponding to the nth case. The input data matrix $\underline{\underline{A}}$ may be retrieved by

$$\underline{\underline{A}} = \underline{\underline{E}}\;\underline{\underline{C}} \;, \tag{13}$$

which exactly replicates each data case in $\underline{\underline{A}}$. One of the primary advantages of EOF analysis arises from the fact that the first few eigenvectors often describe a large portion of the total variance in a sample, depending on the structure and correlation in the field. One may quite accurately approximate the actual field by retaining only the largest few eigenvectors. Assuming 500 cases, the initial data matrix required to describe the synoptic fields is a 120 X 500 matrix, which has 60,000 elements. Using only the first 10 eigenvectors and orthogonal coefficients, the original fields may be represented accurately by multiplication of two matrices, the first a 120 X 10 matrix of truncated eigenvectors, and the second a 10 X 500 coefficient matrix. The total number of elements in both matrices is only 6,200. Since EOF analysis allows a high percentage of the total variation to be explained by only the largest few eigenvectors, it is seen that the data may be accurately estimated using as little as 10% of the total number of data points.

This significant reduction of dimensionality makes EOF's a prime tool to use for climatic estimation, and has been used as such by Horel (1981), Kidson (1975), Walsh and Mostek (1980) and Walsh and Richman (1981) among others.

All N observed fields are represented by the linear combination

$$\underline{a}_n = \sum_{i=1}^{m} c_{in} \, \underline{e}_i \qquad n = 1,2,...,N, \qquad (14)$$

where $\underline{a}$ is the nth cases. Thus each case may be represented as a linear combination of the orthogonal coefficients and elements of the eigenvectors. The first k eigenvectors ($k \ll m$) generally represent a large portion of the total variance in $\underline{a}$. Keeping only the largest k eigenvectors, the actual cases may be very closely approximated by:

$$\underline{a}_n = \sum_{i=1}^{k} c_{in}\, \underline{e}_i \qquad n = 1,2,\ldots,N. \qquad (15)$$

If one retains only significant eigenvectors, maximum information may be retained with little complicating noise. This leads to the obvious problem regarding the optimal number of eigenvectors to keep.

C. SELECTING THE NUMBER OF EIGENVECTORS

In the previous section, it was demonstrated how a data field may be represented accurately by a linear combination of only a small number of eigenvectors and coefficients. The question of how many eigenvectors to retain is vital. Simply stated, the question is at what point does the linear combination no longer add signal, but only describe noise in the data. Unfortunately, there is no single accepted answer to this question. Several possibilities are presented here.

The classical principal component approach is outlined by Morrison (1967), and assumes a very large, normally-distributed sample for the data. In this case, the significant eigenvectors may be identified by asymptotic behavior of the eigenvalues.

One seeks those eigenvectors that are significantly different than zero. Anderson (1963) has shown that sampling problems using normalized data are much more complex than when non-normalized departures from means are used. Therefore, the initial development given here assumes non-normalized data, because the mathematical description is easier to follow. When the number of observations is very large, Anderson (1963) shows the quantity $\sqrt{n}(\ell_i - \lambda_i)$ is distributed normally about a zero mean, with variance of $2\lambda_i$. Here ℓ_i is the sample population eigenvalue, λ_i is the total population eigenvalue, and n the number of cases. Further, Anderson shows the eigenvalues are independent of each other. In this case, one may use a confidence interval approach to determine if the eigenvalues are significantly different than zero. If an eigenvalue is not significantly different than zero, the associated eigenvector describes only random noise. The confidence interval, given by Morrison (1967) is:

$$\frac{\ell_i}{1 + z_{1/2\alpha}\sqrt{2/n}} \leq \lambda_i \leq \frac{\ell_i}{1 - z_{1/2\alpha}\sqrt{2/n}} \qquad (15)$$

where:

$z_{1/2}$ is the standard two tail z score ($z = 1.96$ gives a 95% confidence interval)

The asymptotic decision rule is simply that the eigenvector is discarded unless the lower limit in (15) is greater than zero.

While this method is sound theoretically, and works very well for large data sets, Preisendorfer and Barnett (1977) point out that data sets used in meteorological (and oceanographic) studies are rarely of the size for which asymptotic behavior begins to emerge. In fact, Preisendorfer and Barnett suggest that a sample size on the order of 1000 cases may be required before asymptoticity applies. Since the data set used in this study is much below this size, the classical asymptotic selection approach for determining how many eigenvectors to retain was not used.

Another approach used throughout the literature (e.g., Rinne and Karhila, 1979) involves examination of the natural logarithm of the eigenvalue. This method is called the LEV (Logarithmic EigenValue) diagram method. The basis of this method is that the eigenvectors for those components that describe signal have a different structure than those that describe noise. Furthermore, it has been noticed that the structure change is most easily noted when natural logarithms of the eigenvalues are examined. To use the method, the eigenvalues are first ordered, from largest to smallest. This method will work if there is a distinct change in slope of the ordered eigenvalues at some point. All eigenvalues larger than this slope change point are retained, and all smaller ones omitted. While this method apparently does well in some cases, and is exceedingly simple to use, it is not used in this study for several reasons. First, it is not at all clear that a

break in the slope of the eigenvalues at some point is the demarcation point between those eigenvalues that describe signal and those that describe noise. Secondly, even assuming the break in the eigenvalue slope does indeed mark the point in signal-to-noise domination shift, the method is scientifically unsatisfying because there is little statistical justification for its use.

Another method that appears in the literature is to select the number of eigenvalues and vectors a priori, or select a percent total variance explained value as the cutoff point a priori. Richman (1980) presents several of these methods in detail. For example, Cattell (1958) recommends retaining all eigenvalues necessary to explain 99% of the total variance. Guttman (1954) recommends retention of all eigenvectors associated with eigenvalues larger than 1. Both of these methods in effect involve probable overfactoring. That is, use of these methods leads to keeping more eigenvectors than are actually required to adequately explain the data. This in and of itself is not serious unless the eigenvalues and vectors are rotated to better fit the clusters in space (see Richman, 1981), but it does tend to defeat the purpose of EOF analysis. If overfactoring occurs, one does not receive maximum data reduction. Since the purpose of this study was to reduce the synoptic scale forcing fields to only a few easily separable components to aid in determining typhoon movement, underfactoring is not a real problem.

Richman (1980) used a novel approach to determine how many eigenvectors to retain. He also used rotation of components, which is discussed in detail in the last section of this chapter. His criteria was defined as "meaningfulness". That is, if the component had apparent meaning (if the component field was interpretable synoptically), the component was retained. It has been demonstrated (for example, Craddock and Flood, 1969) that higher order eigenvectors and components degenerate to little more than a series of uncorrelated high and low value regions. This means that there is some scientific justification to Richman's method. Nevertheless, it was not used here because it is entirely subjective, and therefore could give inconsistent results when used by different researchers.

Brown (1981) used the method of retaining the number of components that explain a "reasonable amount" of the total variance. Specifically, using the same grid and data fields that are used in this study, he carried out experiments in map typing using the largest 10, 15 and 20 of the 120 eigenvectors. This selection approach is rather arbitrary, since there is no objective way of distinguishing what the eigenvectors are representing with respect to the signal-noise problem, and specifically, if any signal is being omitted.

The final method, which is used in this study, is based on a selection method introduced by Preisendorfer and Barnett (1977). In essence, the scheme is a Monte Carlo approach to determining the number of eigenvectors to keep. It is not

very different from the classic asymptotic appraoch described by Morrison (1967). The main difference is that it is assumed by Preisendorfer and Barnett that not enough cases are available to use an asymptotic approach with geophysical data bases. One key assumption is that the true (physical) variables are normally distributed at all individual grid points. The simulation input data are normally distributed, with mean zero, variance one, which is just simulation of point normalized data. Given these constraints, and using a large number ($N \geq 100$ is recommended by Preisendorfer and Barnett (1977)) of simulations, one can create sufficient numbers of random fields to simulate accurately the eigenvalues that result if the process is purely random. In addition to calculating the mean value of the simulated eigenvalue, the standard deviation of that eigenvalue is calculated over the 100 or more simulations. If the true physical eigenvalues deviate from the simulated random field eigenvalues by more than two (three) standard deviations, one is 95% (99%) confident that the field is significantly different from a field that is purely random. In other words, if deviation is by more than two standard deviations, one is reasonably assured that the eigenvector is describing signal rather than noise. The simulated eigenvalues obtained in this study will be presented in the next chapter, along with the eigenvalues obtained from analysis of the data. In using this Monte Carlo method, 504 simulated 120 point random grids were obtained. The eigenvalues of these random fields were found and stored. This process was repeated 100 times to obtain the

simulated eigenvalues and standard deviations of the eigenvalues. These were then compared to the true data eigenvectors. One caution must be stated concerning use of this method. Richman (1980) points out that this method has potential to slightly underfactor. However, this is not of primary concern here since the potential for underfactoring is only slight.

D. ROTATION OF VECTORS

Rotation methods seek to rotate the eigenvectors (axes) in space to better fit data clusters. There is some controversy existing (Richman, 1980) as to whether rotation of the resultant components (eigenvectors) should be employed. Many of the potential schemes have been surveyed in detail by Richman (1980), who describes some of the specific strengths and weaknesses of the schemes.

A very simple example of rotation follows. Suppose that two distinct data clusters are positioned (in Cartesian two-dimensional space) at $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$. Following the method outlined earlier in this chapter, the eigenvalues would then be $\begin{bmatrix}4.5\\.5\end{bmatrix}$ (for non-normalized input data). The eigenvectors would be $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\-1\end{bmatrix}$ respectively. It is noted then the first eigenvector (which explains 90% of the total variance) bisects the two data clusters in space. The second eigenvector does not really fit the data clusters. Even the first eigenvector does not give a true representation of the clusters in space. Misrepresentation of this type may be eased by use of rotation. The two broad classes of rotation that are employed are the

orthogonal and the oblique. Orthogonal rotation pivots the eigenvectors identically so as to maintain the orthogonal relationship. It is seen in the simplified case just presented that an orthogonal rotation would never give a perfect representation of the input clusters, as the input clusters only have a 45° angle between them in the two dimensions, and are assumed to occur with equal frequency. Oblique rotation, on the other hand, pivots the vectors so as to most closely fit the data clusters without necessarily retaining the orthogonality constraint. In the simplified case just presented, the vectors would be pivoted (within a scaling factor) to $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$. The vectors are no longer orthogonal, nor is it possible to determine quantitatively the amount of total variation explained by either of the vectors without exhaustive analysis. Richman (1981) uses pre-determined input fields to simulate the principal component processes. He then compares non-rotated components to both orthogonally and obliquely rotated components. His results show obliquely rotated components give vastly improved delineation of the input patterns. He then concludes that obliquely rotated components are a better tool to use for map typing than either orthogonally rotated or non-rotated components. If the purpose is to identify and interpret all types of meteorological patterns that force another event, obliquely rotated components would appear to give superior results.

Rotation was not used in this study for several reasons. Delineation of patterns of meteorological features was not the

specific purpose of this research. EOF's were used in this study for two purposes. First, they were used to obtain the orthogonal coefficients which are used in the formulation of regression equations to forecast tropical storm movement. Secondly, they were used to reduce the data. The first purpose of the research makes physical identification and interpretation of the resultant eigenvalues less critical. It is the orthogonal coefficients derived from the linear combination of the eigenvectors that are used, not the actual eigenvectors themselves. Nevertheless, it is desirable to use the resultant eigenvectors with certainty to identify and interpret the forcing features. It is primarily due to the data reduction purpose of this study that use of rotated components becomes less attractive. Since the amount of explained variance (by each component) is unknown after rotation, the question of how many eigenvectors to retain becomes unclear. In fact, perhaps the only valid criteria for retention becomes Richman's meaningfulness criteria. In any case, the problem of determining how many vectors to retain becomes much more difficult after rotation has been employed.

An even more insidious problem with rotation of the vectors is the effect of underfactoring on the resultant vectors. Richman (1981) also experiments with underfactoring. If too few vectors are retained and rotated, then the resultant rotated vectors become combinations of vectors associated with several actual input data clusters. Therefore, if underfactoring

exists, the same type of bisection that is seen in the worst possible case with unrotated vectors may occur with the rotated vectors. Since data reduction in this study is paramount, rotation of components seems ill-advised at the present time.

As a final note, Richman's results, and the simplified results shown at the beginning of this section clearly show non-rotated components may not represent the true synoptic patterns. Conceptually, if the data clusters (input data) are not symmetric, errors in the EOF representation are less likely. This is perhaps most easily seen with a simplified example. If, for instance, in two dimensions, there are two data clusters occurring with equal frequency, one of the resultant eigenvectors will bisect the two clusters. This is the case in the simplified example above since the two cluster points were assumed to occur with equal frequency. If the clusters do not occur equally, this bisection does not occur. Richman's simulated fields were input in mirror-image pairs, with equal probability of occurrence. In this case, the resultant eigenvector bisected the given input fields. True geophysical synoptic fields are not orthogonal in nature (Barry and Perry, 1973 and others). On the other hand, it is anticipated that true geophysical fields do not come in matched opposite pairs that occur with similar frequency. It is for this reason that the first several unrotated vectors should indeed represent actual synoptic variability patterns.

IV. RESULTANT EMPIRICAL ORTHOGONAL FUNCTIONS

The mathematical and theoretical framework for EOF analysis was developed in Chapter III. In this chapter, the forcing of each eigenvector on tropical storm movement is examined by correlation of storm motion with the strength of the particular vector for a given data case, which is given by the value of the orthogonal coefficient associated with the vector. Before any meaningful analysis of physical forcing on typhoon motion may be attempted, the actual eigenvectors must be examined.

Following the mathematical development of Chapter III, the 120 X 504 data matrix was normalized at each grid point, and the eigenvectors were obtained for all three data levels (500, 700 and 850mb). The resultant eigenvalues for all three levels were then compared to the random eigenvalues generated from Monte Carlo simulation using 100 simulations, as suggested by Preisendorfer and Barnett (1977). These Monte Carlo eigenvalues were all computed from 120 X 504 matrices whose elements were random normal variables with a mean value of zero and a standard deviation of one. Thus the statistical structure of the random fields is identical to the real data normalized fields. The value of the eigenvalues for the three levels is given in Table 4-1, which also gives the cumulative percent explained total variance for each successive eigenvector. Table 4-2 is a list of the randomly generated eigenvalues and their standard deviations for comparable modes. If the real data

TABLE 4-1

Eigenvalues and cumulative percent explained variance (in parenthesis) for the normalized D-value at each level.

EIGENVALUE	500MB	700MB	850MB
1	39.863 (33.3)	29.569 (24.7)	40.143 (33.5)
2	21.597 (51.3)	25.950 (46.4)	20.387 (50.5)
3	9.287 (59.1)	11.139 (55.7)	11.093 (59.8)
4	7.430 (65.3)	7.859 (62.2)	8.380 (66.8)
5	6.083 (70.4)	7.107 (68.2)	6.084 (71.9)
6	5.293 (74.8)	5.103 (72.4)	4.137 (75.3)
7	3.990 (78.1)	3.777 (75.6)	3.428 (78.2)
8	3.130 (80.7)	3.622 (78.6)	2.931 (80.7)
9	2.490 (82.8)	2.596 (80.8)	2.591 (82.8)
10	2.111 (84.6)	2.291 (82.7)	2.153 (84.6)
11	1.852 (86.1)	1.956 (84.3)	1.743 (86.1)
12	1.701 (87.5)	1.607 (85.7)	1.519 (87.3)
13	1.409 (88.7)	1.416 (86.9)	1.279 (88.4)
14	1.202 (89.7)	1.256 (87.9)	1.059 (89.3)
15	1.030 (90.6)	1.128 (88.9)	.949 (90.1)
16	.946 (91.4)	1.064 (89.7)	.888 (90.8)
17	.787 (92.0)	.971 (90.6)	.841 (91.5)
18	.776 (92.7)	.774 (91.7)	.657 (92.1)
19	.690 (93.3)	.623 (92.2)	.562 (92.5)
20	.581 (93.7)	.616 (92.7)	.552 (93.0)
⋮	⋮	⋮	⋮
40	.111 (98.2)	.152 (97.3)	.134 (97.6)
⋮	⋮	⋮	⋮
60	.039 (99.4)	.060 (98.8)	.055 (98.9)
⋮	⋮	⋮	⋮
80	.016 (99.8)	.029 (99.5)	.027 (99.5)
⋮	⋮	⋮	⋮
120	.000 (100.0)	.001 (100.0)	.002 (100.0)

TABLE 4-2

Eigenvalues and standard deviations corresponding to the modes in Table 4-1 as generated by the Monte Carlo method (see description in text).

MODE	EIGENVALUE	STANDARD DEVIATION	EIGENVALUE PLUS TWICE STANDARD DEVIATION
1	2.169	.044	2.258
2	2.100	.037	2.174
3	2.048	.031	2.110
4	2.005	.030	2.065
5	1.964	.025	2.018
6	1.928	.026	1.981
7	1.894	.025	1.944
8	1.862	.023	1.909
9	1.831	.023	1.879
10	1.802	.023	1.854
11	1.775	.022	1.819
12	1.749	.020	1.790
13	1.725	.021	1.766
14	1.699	.019	1.737
15	1.675	.019	1.713
16	1.652	.021	1.694
17	1.628	.018	1.664
18	1.604	.018	1.639
19	1.581	.017	1.614
20	1.538	.019	1.595
..	..	..	..
..	..	..	..
..	..	..	..
40	1.203	.014	1.231
..	..	..	..
..	..	..	..
..	..	..	..
60	.926	.011	.948
..	..	..	..
..	..	..	..
..	..	..	..
120	.273	.010	.293

eigenvalue for a specific mode is greater than the random eigenvalue plus twice the standard deviation, the eigenvalue and corresponding eigenvector represent geophysical signal, and the eigenvector is retained. To facilitate this comparison, the value of the random eigenvalue plus twice the standard deviation is also given in Table 4-2. The values of the standard deviations in Table 4-2 are consistent with Preisendorfer and Barnett's (1977) results. Comparisons of the three actual field eigenvalues to those of the random field are conducted separately, since the number of significant eigenvectors may be different for each level. The only relationship between the eigenvectors of the three levels comes from any dynamic vertical coupling that may exist.

Several interesting features emerge from examination of the eigenvalues. The number of eigenvectors to retain is different depending on the retention scheme chosen. For example, Guttman's lower bound test suggests retention of the first 14 or 15 eigenvalues for these levels. Cattell's 99% retention rule would indicate retention of more than 40 modes at each level. The Preisendorfer and Barnett selection scheme is much less conservative, and suggests retention of only 10 eigenvectors at 850 and 500mb and 11 at 700mb. Because the Preisendorfer and Barnett method keeps fewer modes, the potential for underfactoring increases. Since only 10 or 11 eigenvectors are to be retained, roughly 15% of the variance in the fields is directly accountable to random fluctuations (noise). This amount of unexplained variance is not unrealistic in the

tropics. These errors are most likely due to either intialization or measurement error in the fields. This is not surprising because the initialization problem in the tropics is difficult (weak governing mass-wind balance relationship). Even more importantly, there is a very small gradient in the geopotential field, except in the region near the tropical storm. This would tend to give a greater weighting to any observational error in the tropics, compared to the mid-latitudes, where a linear balance initialization with quasi-geostrophic constraints can be imposed to reduce errors in the height fields. Since the areal extent of the grid incorporates a large portion of the tropical synoptic forcing field (Fig. 2-1) it is entirely conceivable that there is a 15% level of random error in the D-value fields.

The 500mb eigenvalues from Table 4-1 are graphically compared to the Monte Carlo simulated eigenvalues (Table 4-2) in Fig. 4-1. It is seen the actual 500mb eigenvalues decrease very rapidly with increasing mode, which indicates that a large number of the components represent data clusters containing random noise. Graphs of the 700 and 850mb eigenvalues are not included because they are very similar to the 500mb values.

Preisendorfer and Barnett's assertion that asymptoticity does not apply for a sample size of 504 data cases may also be examined. If the asymptotic results are valid, the ratio

$$\frac{\bar{\ell}_i}{s_i} = \left[\frac{(n-1)m}{2}\right]^{1/2} = 174$$

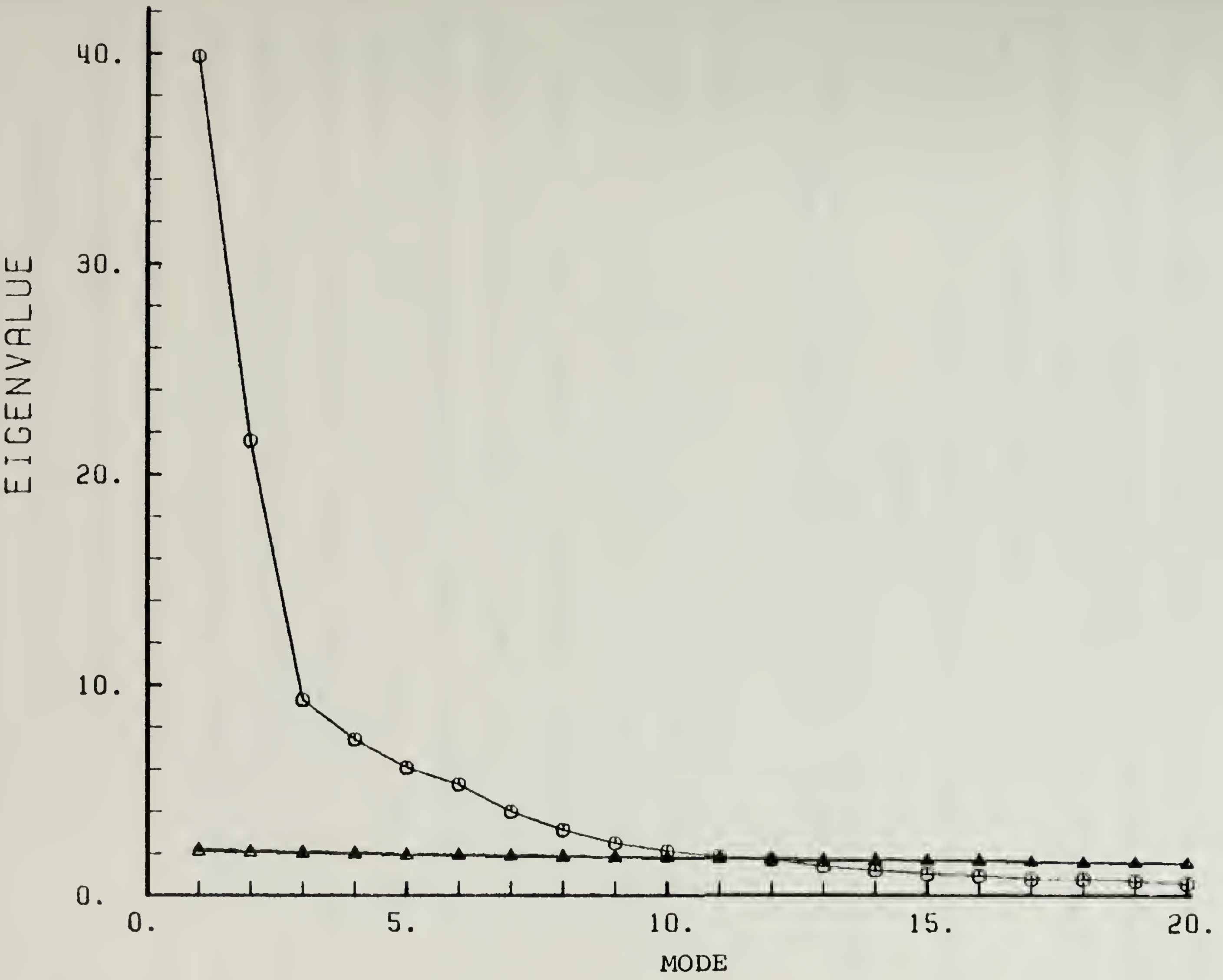


Fig. 4-1. The largest twenty true eigenvalues of the 500mb D-value fields compared to the Monte Carlo generated eigenvalues for the same modes. Monte Carlo eigenvalues are denoted by a triangle, the true 500mb values by a circle.

should be very nearly constant. Here $\bar{\ell}_i$ is the mean randomly generated ith eigenvalue, s_i is the standard deviation for the ith mode, n is the number of cases and m is the number of grid points. The value of this ratio is given in Table 4-3 for selected modes. It is seen that the ratio is not constant, nor does it approach the theoretical value expected for asymptoticity. Thus it is concluded that asymptotic theory is not valid for this study.

TABLE 4-3

Test parameter for the asymptotic theory of eigenvalues is shown for various modes (see text for details).

MODE	1	2	5	10	15	20	40	60	120
RATIO	49.3	56.8	78.6	78.6	88.2	80.9	85.9	84.2	27.3

Based on these tests for significant eigenvectors, it was decided to retain the largest 10 eigenvectors for all levels. These first 10 eigenvectors at 500mb are shown in Figs. 4-2 through 4-11 and will be examined in detail. The first 10 eigenvectors for both the 700 and 850mb level are shown in Appendix A, without comment. The discussion of the first 10 eigenvectors at 500mb will include an interpretation of the probable forcing that the particular pattern has on the tropical storm, which is always at grid point 70.

The actual values of the eigenvectors in Figs. 4-2 through 4-11 are non-dimensional, since normalized data are used on input. The broad scale forcing features of an eigenvector do

have meaning in the standard meteorological sense. Areas of higher values of the eigenvector may properly be thought of as high pressure (D-value) regions, areas of low elements as low pressure regions, and more strongly packed isopleths indicate stronger flow regions. Finally, it is stressed that each eigenvector actually represents the pattern shown and the exact inverse of the pattern shown. Relative gradients of the patterns and positions of the closed isopleth features remain unchanged for the positive or inverse eigenvectors. All following discussion will be made using the eigenvector pattern shown; the inverse case will not be discussed. Relevant features for the inverse pattern may easily be obtained following the same reasoning as below.

Eigenvector 1 (Fig. 4-2): This pattern shows a band of stronger easterlies directly to the north of the cyclone. Additionally, there is a slight northerly component to the flow directly upstream of the storm. The forcing of the tropical cyclone for this type of pattern should be to the west and south.

Eigenvector 2 (Fig. 4-3): This component shows small gradients throughout the field, as expected in the tropics. As with pattern 1, a broad band of easterlies is seen to the north of the storm, but they are much farther north than for pattern 1. A primary difference between this component and the first vector is that there appears to be a low centered south-southwest of the storm, while this low was to the south-southeast for

vector 1. This component and component 1 both exhibit properties of planetary scale waves, as they both have very low wavenumber over the 70 degree longitudinal span of the chart. This pattern should induce weak forcing to the west and to the south.

Eigenvector 3 (Fig. 4-4): An entirely different type of pattern compared to the first two components is seen here. The vector has a fairly strong area of lower values to the west, with a small higher valued area south-southeast of the storm. Another small low is seen well to the northeast corner of the pattern. Forcing on the storm should be to the north (strongly) and east (weakly).

Eigenvector 4 (Fig. 4-5): The predominant feature of this vector is a well developed low to the north and east of the storm. The storm itself appears to be situated in a strong flow region between a high and low. The forced motion should be strongly to the east, with a weak drift to the south.

Eigenvector 5 (Fig. 4-6): A strong high valued area directly to the north of the storm is the predominant feature in this eigenvector. The pattern is essentially weavenumber 1 across the 70 degree span of the chart. The physical analogue of this vector is difficult to determine. It could well be that this is a bisection of two distinct data clusters of high pressure on the outer extremities of the grid, since this pattern bears strong resemblence to the non-rotated bisection case simulated by Richman (1981). In any case, the eigenvector is

usable with coefficients that appear in the formulation of regression equations, and does indeed describe a global wavenumber 5 pattern. This pattern should force tropical storms to the west and north.

Eigenvector 6 (Fig. 4-7): This pattern is another wavenumber 1 across the 70 degree longitude span of the grid (global wavenumber 5). The dual low centers are generally similar to the pattern in eigenvector 3. The forced motion of the tropical cyclone should be to the west, with little meridional forcing.

Eigenvector 7 (Fig. 4-8): The expected higher degree of complexity for higher order modes is beginning to show in this vector. Five well-defined high or low centers are seen in the pattern. This vector is approximately global wavenumber 7, so that with this eigenvector the slow transition from large scale to smaller synoptic scales is beginning. The physical meaning of the pattern is also becoming more difficult to define. The forcing of the storm should be weakly to the north and west.

Eigenvector 8 (Fig. 4-9): As with eigenvector 7, there is a complex pattern of well-defined high and low value centers, with the storm located in the northern regions of a high center. Forcing to the east and south is anticipated from this pattern, although all forced motions should be weak.

Eigenvector 9 (Fig. 4-10): Eigenvector 9 is somewhat surprising since it has less complexity than the preceeding two

eigenvectors. Nevertheless, it is approximately global wavenumber 7. A strong blocking high center is found directly to the west of the storm, while the storm itself is on the west side of a weaker low. It is possible that the blocking high pattern represents the effect of the 500mb anticyclone east of the Tibetan Plateau heat low. Motions forced from this pattern should be weakly to the south and east.

Eigenvector 10 (Fig. 4-11): The final eigenvector retained in the truncated set of 10 is the most complex. A series of well developed highs and lows are seen throughout the extent of the grid. Short range forcing on the storm would come from a high located south of the cyclone and two strong low centers flanking the storm. The pattern is wavenumber 2 over the 70 degrees covered by the grid and corresponds to a global wavenumber 10. This pattern defines even smaller synoptic scale forcing than the previous patterns. Perhaps coincidentally, the eigenvector 10 for the 700mb data set (Appendix A) is virtually identical. This similarity indicates this pattern is probably a true physical signal, which is vertically coupled through the mid-troposphere. Motion forced from this pattern will be to the south with little zonal forcing.

It is essential to show how these ten eigenvectors just described would combine to represent the original field. Selection of a case on 0000GMT 27 August 1967 was made at random to demonstrate the reconstruction. At this time, Typhoon Marge was located at approximately 18°N 125°E with maximum winds of 125 knots. The actual 500mb D-value field is shown in

Fig. 4-12. The areal extent of the grid is from 43° to 8°N, and 85° to 155°E. Therefore, this grid encompasses both tropical and mid-latitude forcing on the storm. A linear combination of the first ten eigenvectors and the associated orthogonal coefficients should be adequate to represent the relevant physical features according to the discussion in Chapter III.B.

Among the salient features seen in the total field (Fig. 4-12) is a strong blocking high pressure to the northwest of the typhoon, positioned at about 25°N, 100°E. A 500mb high pressure at this location is east of the Tibetan Plateau heat low which is a stationary feature of the planetary circulation. There is also a strong high pressure cell (D-values in excess of +320 meters) to the northeast of the typhoon. This second high pressure is the westward extension of the subtropical anticyclone over the western Pacific. Well to the north of the cyclone is a strong band of mid-latitude westerlies. A well-developed trough extends from the westerlies into the tropics and encircles the typhoon.

As the input data have been normalized, the fields need to be reconstructed using

$$d_i = \sum_{n=1}^{m} (c_i \underline{e}_{in}) s_i + \bar{d}_i, \quad i = 1,2,\ldots,120,$$

where m is the number of eigenvectors and orthogonal coefficients used in the reconstruction, $\bar{d}_i$ and s_i are the mean

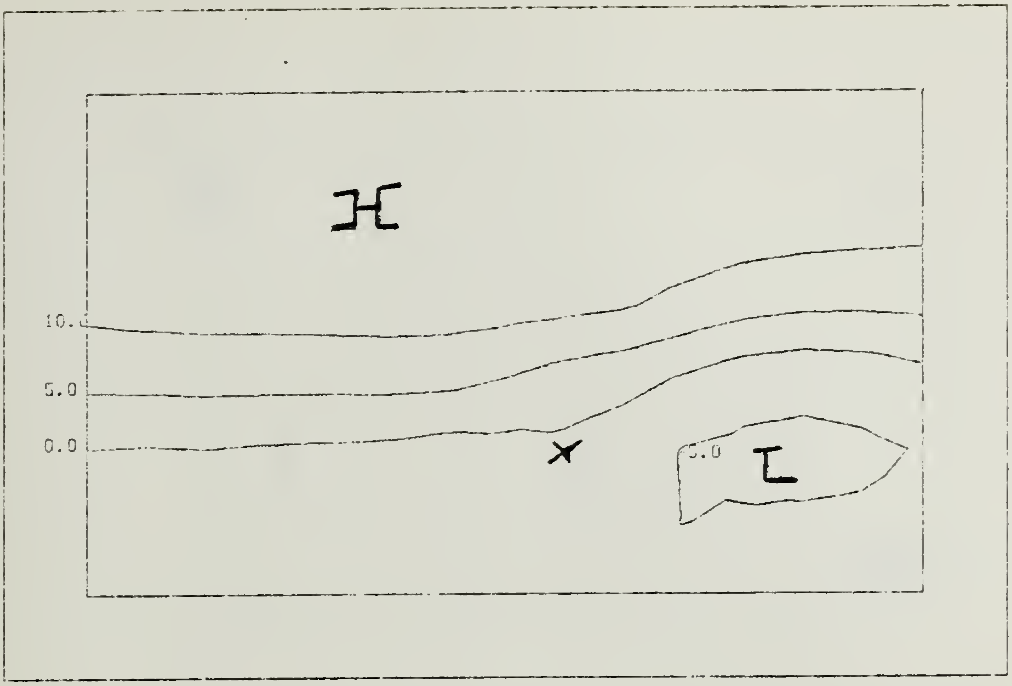


Fig. 4-2. Eigenvector 1 elements (multiplied by 100) at 500mb with the tropical cyclone located at the x-position.

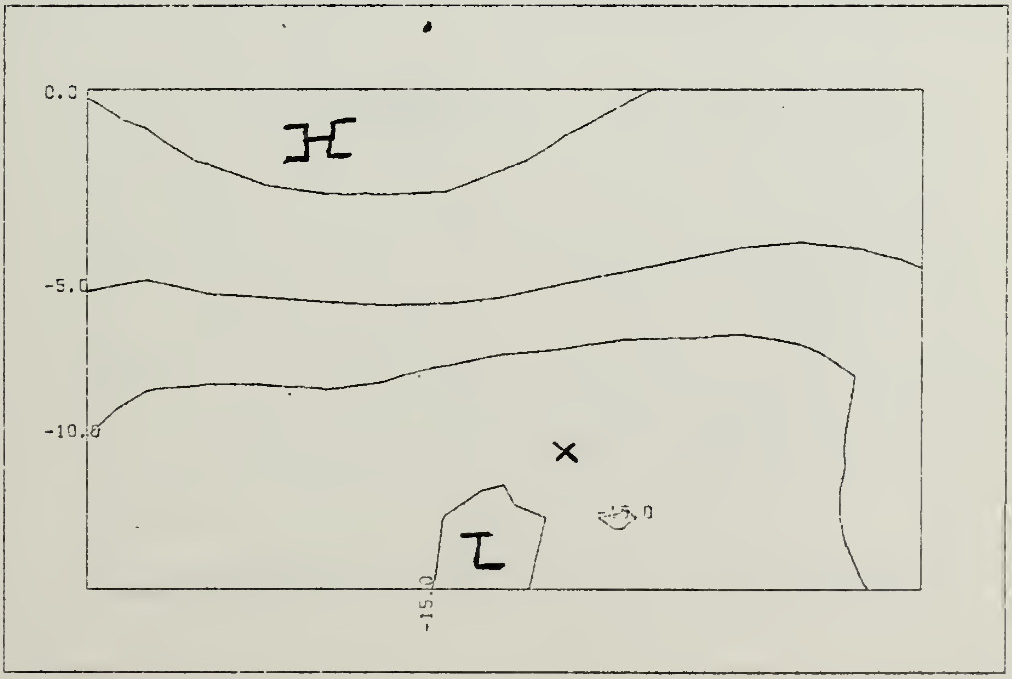


Fig. 4-3. Similar to Fig. 4-2 except for eigenvector 2.

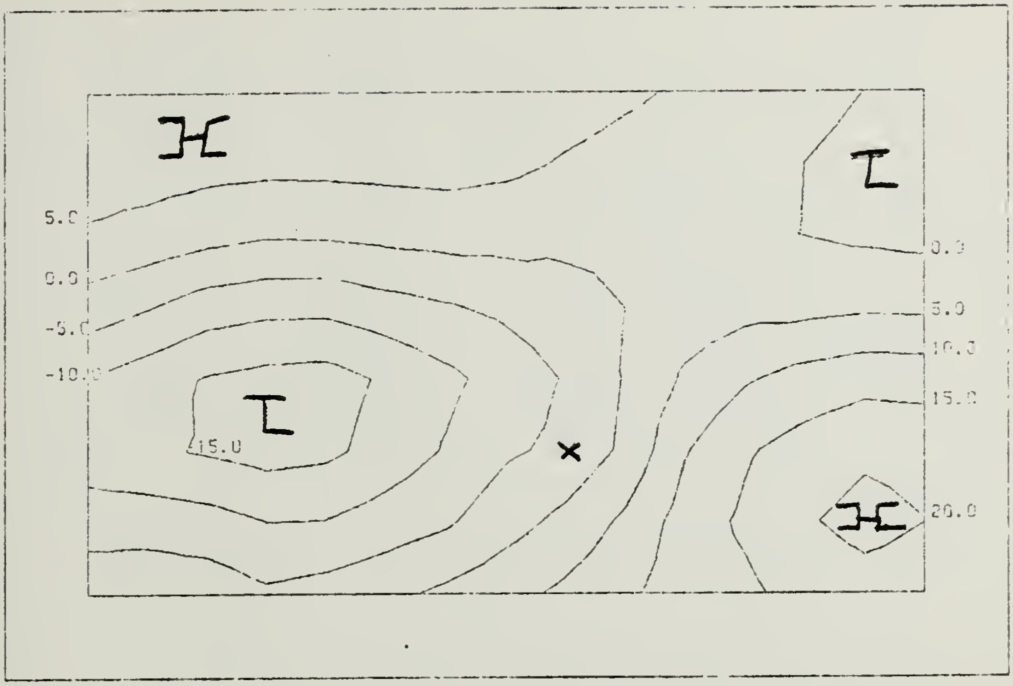


Fig. 4-4. Similar to Fig. 4-2 except for eigenvector 3.

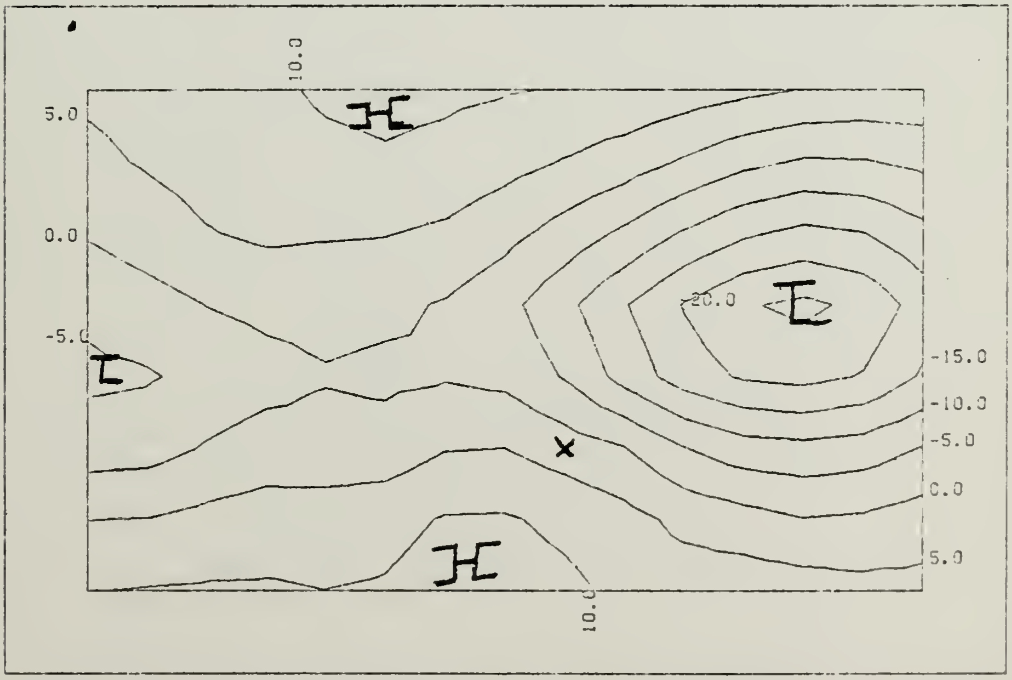


Fig. 4-5. Similar to Fig. 4-2 except for eigenvector 4.

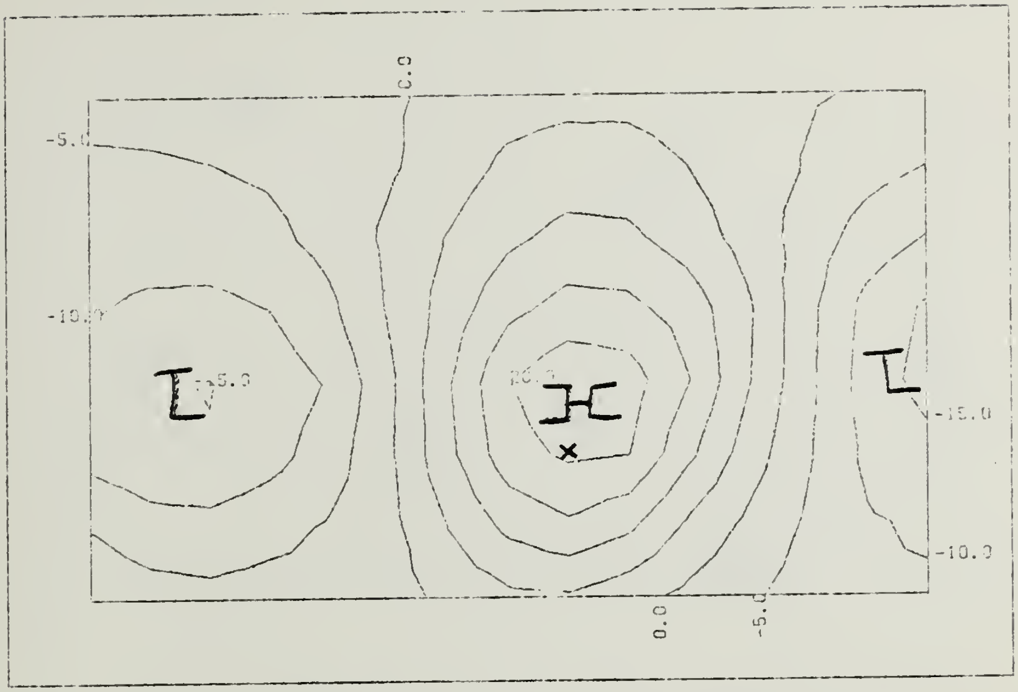


Fig. 4-6. Similar to Fig. 4-2 except for eigenvector 5.

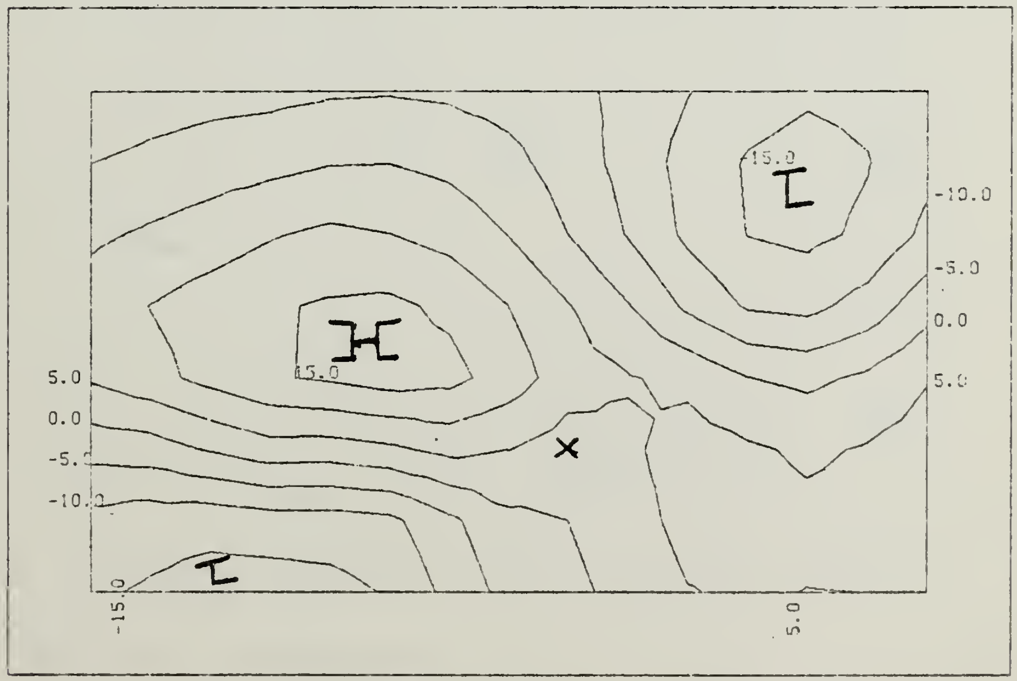


Fig. 4-7. Similar to Fig. 4-2 except for eigenvector 6.

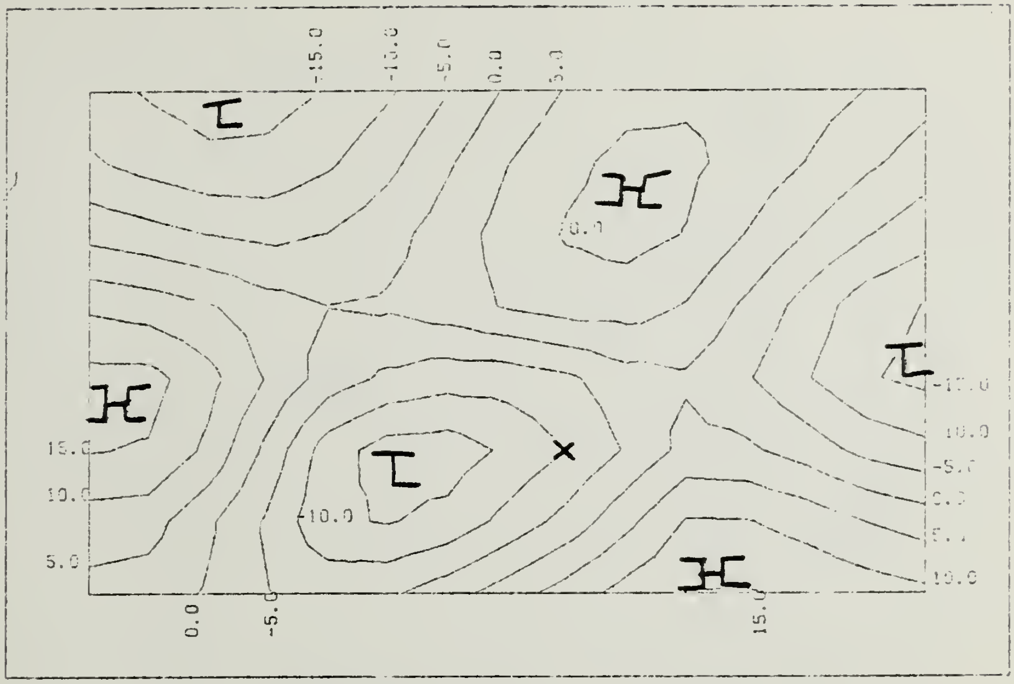


Fig. 4-8. Similar to Fig. 4-2 except for eigenvector 7.

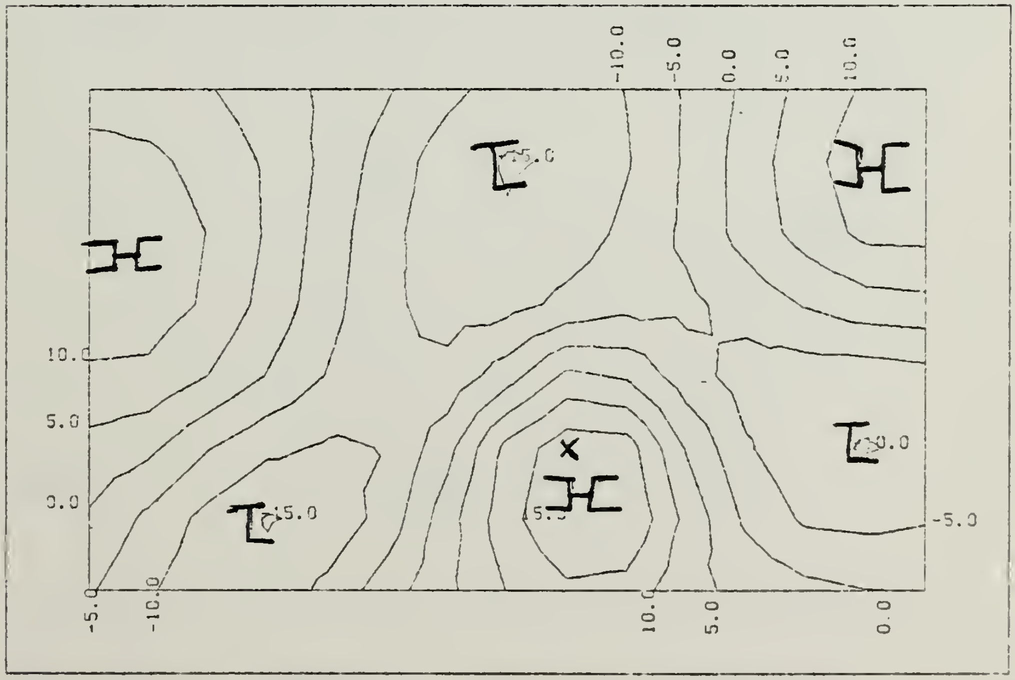


Fig. 4-9. Similar to Fig. 4-2 except for eigenvector 8.

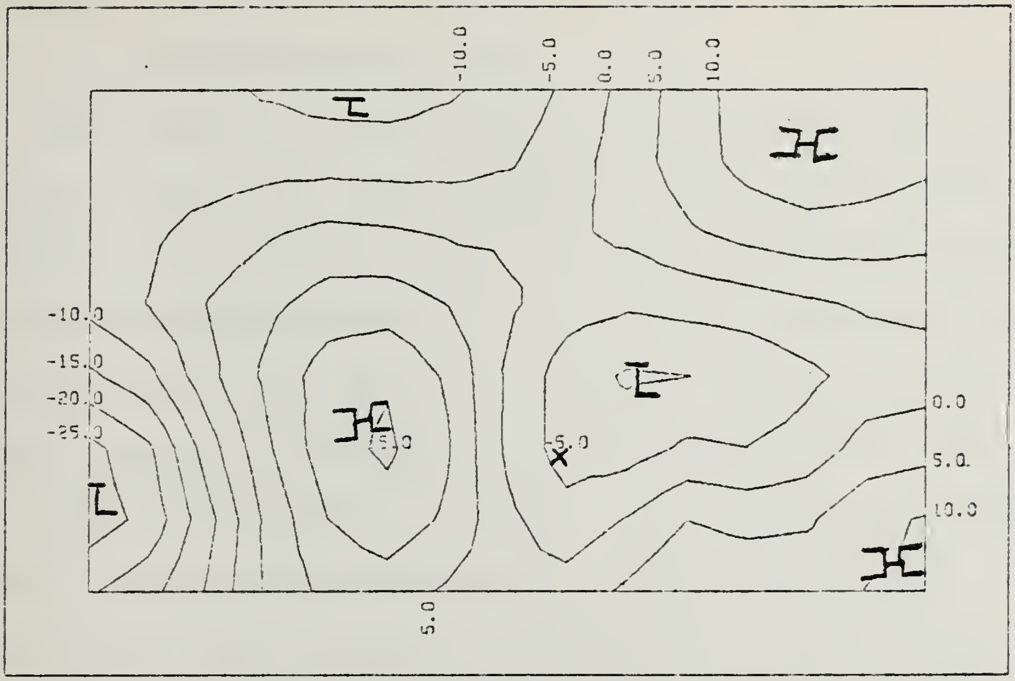


Fig. 4-10. Similar to Fig. 4-2 except for eigenvector 9.

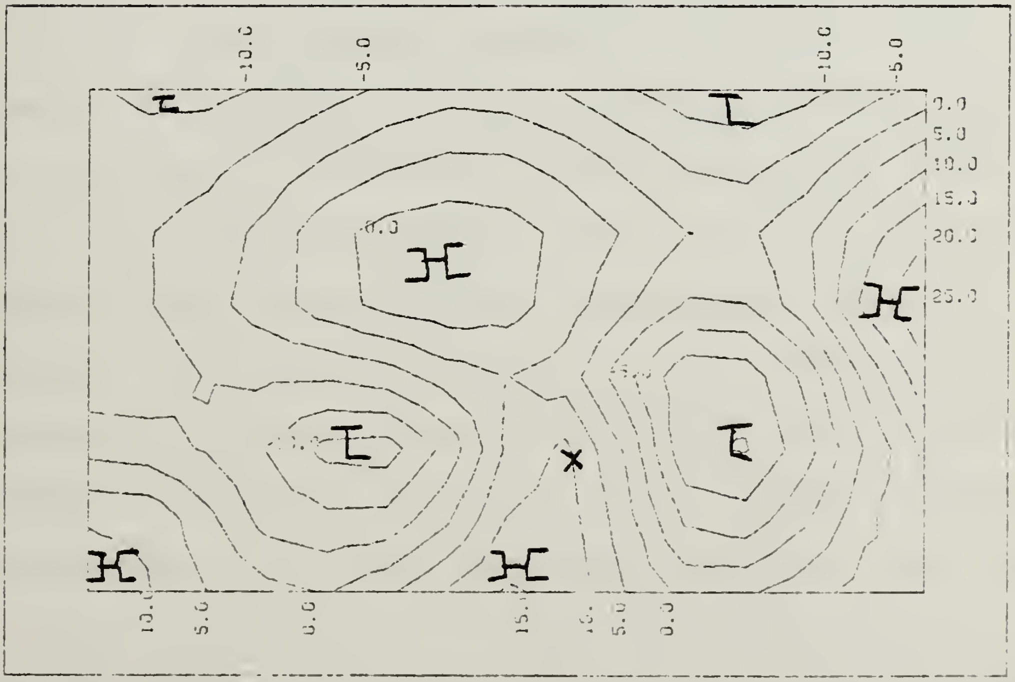


Fig. 4-11. Similar to Fig. 4-2 except for eigenvector 10.

and standard deviation of the D-value at the ith grid point, and d_i is the reconstructed value.

The reconstructed field using only the first vector and coefficient (Fig. 4-13) shows westerlies well to the north with a ridge circling over the top of the storm from the east. The general features revealed by use of this eigenvector are the westerlies and high to the northwest. When the second and third vectors are included in the reconstruction (Fig. 4-14), little information is gained. This is expected since these two patterns are not evident in the actual field.

The inverse of the fourth eigenvector has similarities to the actual case being reconstructed. Both patterns show a high pressure to the northeast and northwest of the storm with a trough in the northern section of the grid. It is anticipated that addition of this eigenvector should greatly improve resolution of features on the reconstructed field. Changes in the field are evident on Fig. 4-15, but the overall resolution of the features is not dramatically improved. Nevertheless, inclusion of this vector does increase the high pressure cell to the northeast of the typhoon, and increases the gradient between the mid-latitude and tropical regions.

The inverse of the fifth eigenvector also has many similarities to the original field. A significant improvement in the shape of the general features is seen after the fifth vector is added (Fig. 4-16). A slight trough appears in the mid-latitude westerlies and a coupling of the tropical and mid-latitude trough is seen for the first time. Inclusion of the

next three eigenvectors (vectors 6 through 8) add very little to the reconstructed field, and are not shown. Similarities between eigenvector 9 and the original field include a sharp trough in the westerlies which connects with a tropical trough in the vicinity of the typhoon. When this eigenvector is added to the linear combination of the previous eight, the broad scale pattern (Fig. 4-17) is delineated much better. There is general agreement in the positions of the large-scale features and the gradients between them. Further refinement through use of higher order modes is necessary to obtain the actual chart. The difference between the patterns in Fig. 4-12 and 4-18 is, according to the analysis here, simply random noise. Nevertheless, with only the first nine eigenvectors the salient features have emerged, and major forcing from the large scale on the typhoon is defined. The continued progression in the reconstructed fields using 10, 20 and 40 eigenvectors are shown in Figs. 4-18 to 4-20. It is noted that the reconstructed field is almost exact after 40 terms are included, and some features due to random noise in the field are reproduced. The correlation of the reconstructed field using various modes to the original field is shown in Table 4-4. It is seen here that the correlation of the two fields asymptotically approaches 1 as the number of modes in the reconstruction is increased. Furthermore, large jumps in the correlation are seen when the first and ninth eigenvectors are added, and smaller jumps are seen with inclusion of the third and fourth vectors. This is

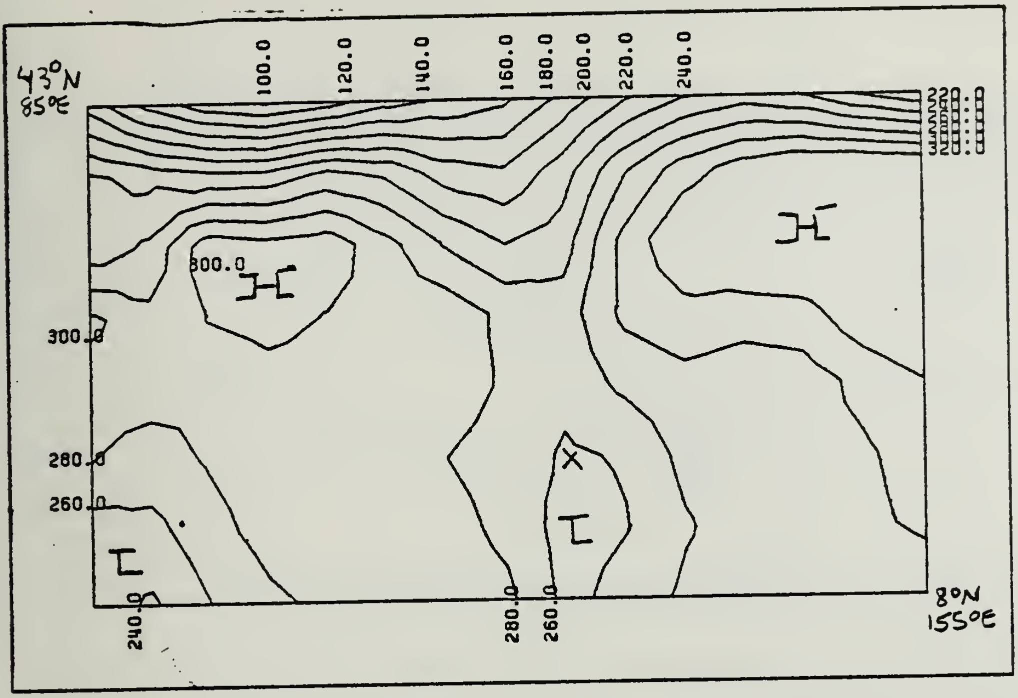


Fig. 4-12. 500mb D-value (meters) field surrounding Typhoon Marge at 0000GMT 27 August 1967. Marge is located at 18°N 125°E (location X).

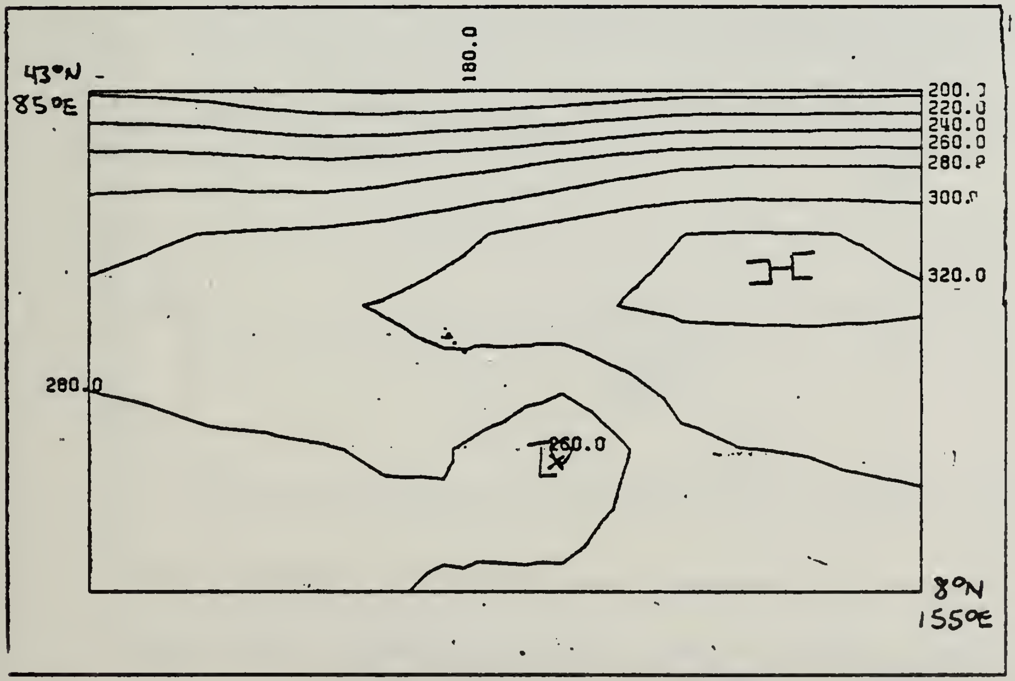


Fig. 4-13. Reconstruction of 500mb D-value field, 0000GMT 27 August 1967, using the first eigenvector and orthogonal coefficient. This compares to true field (Fig. 4-12).

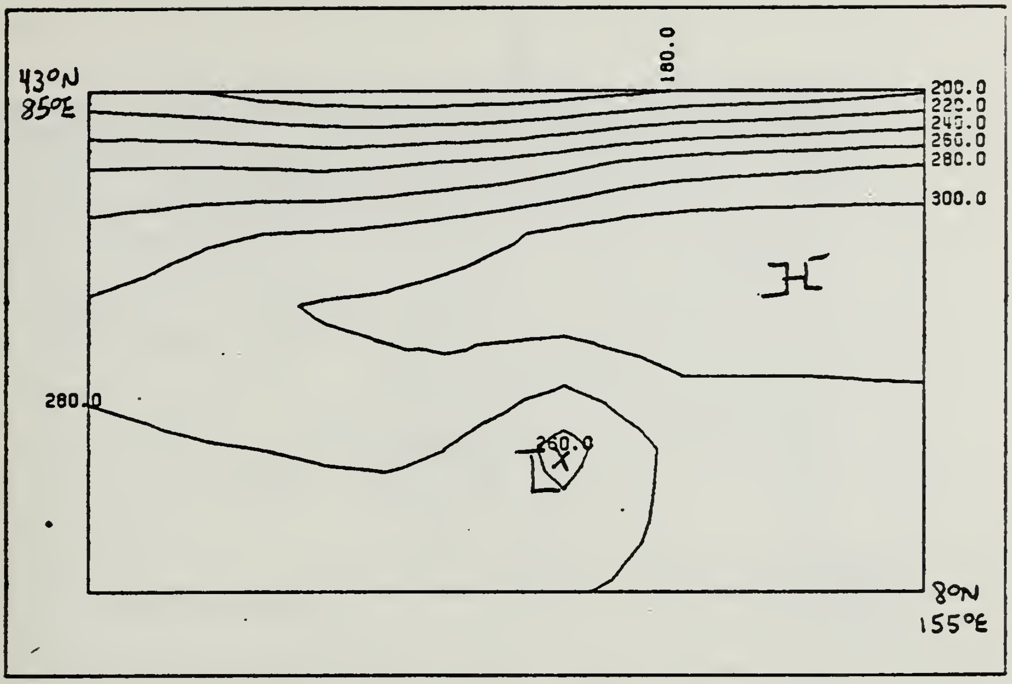


Fig. 4-14. Similar to Fig. 4-13, except first three eigenvectors are used in reconstruction.

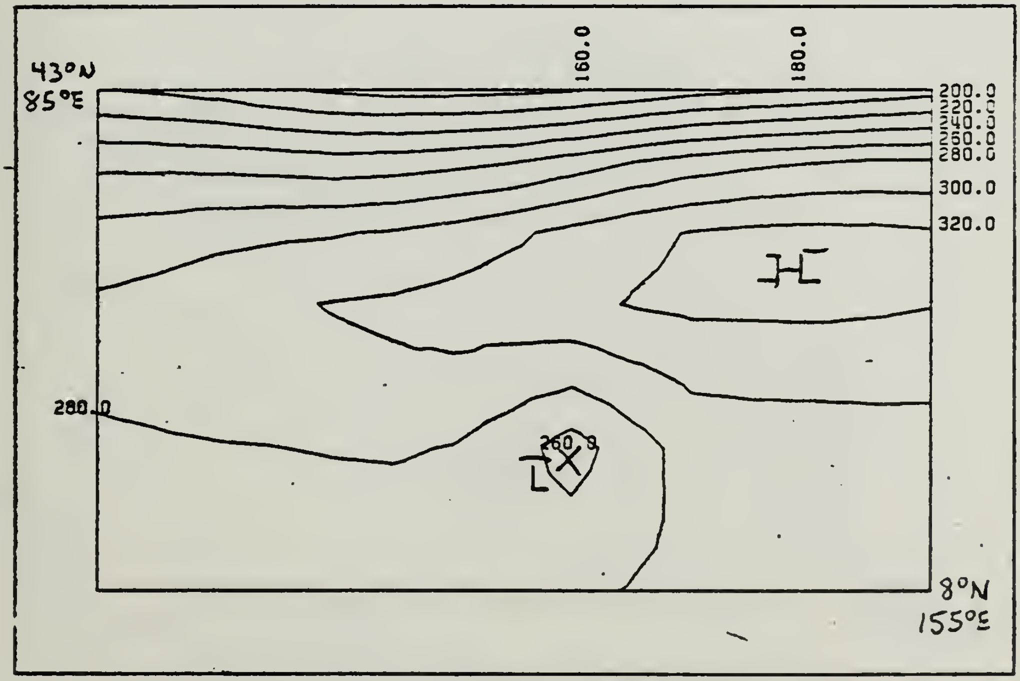


Fig. 4-15. Similar to Fig. 4-13, except first four eigenvectors are used in reconstruction.

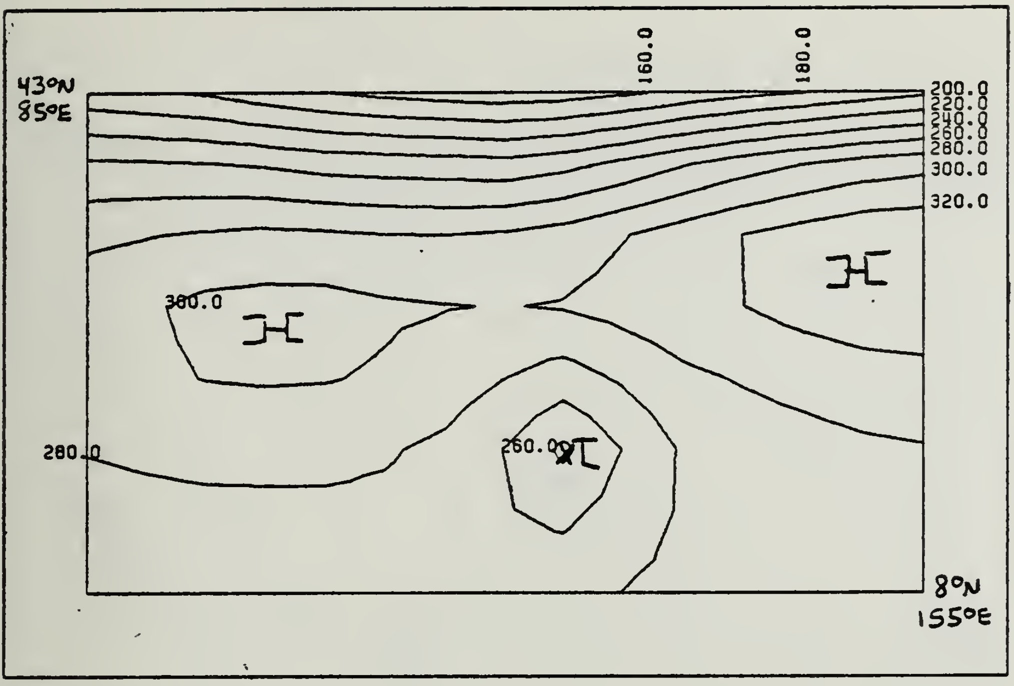


Fig. 4-16. Similar to Fig. 4-13, except first five eigenvectors are used in reconstruction.

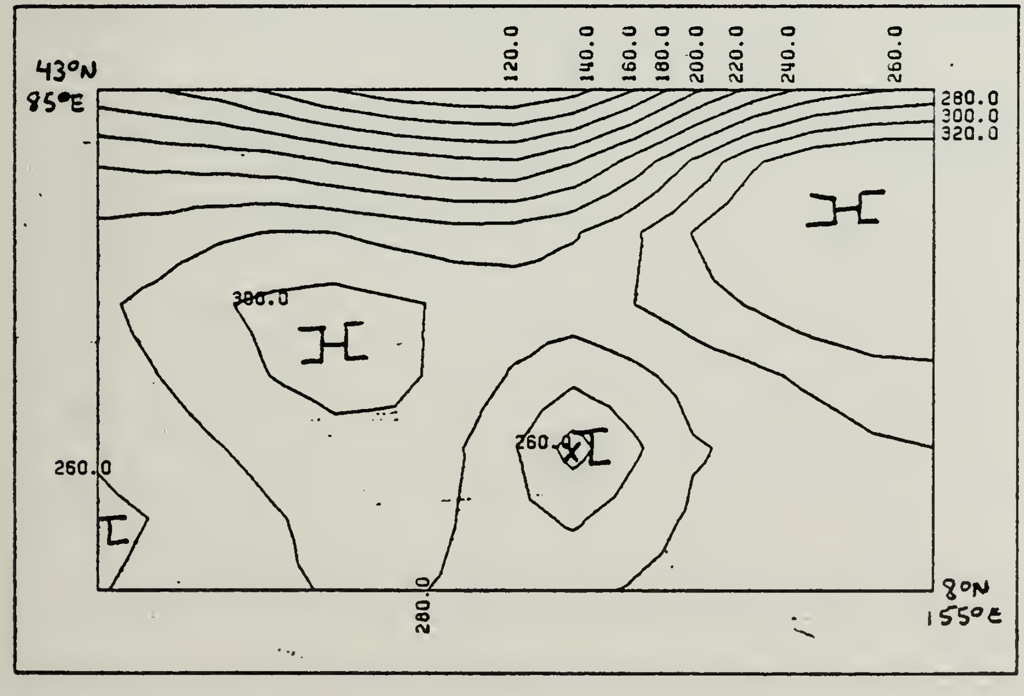


Fig. 4-17. Similar to Fig. 4-13, except first nine eigenvectors are used in reconstruction.

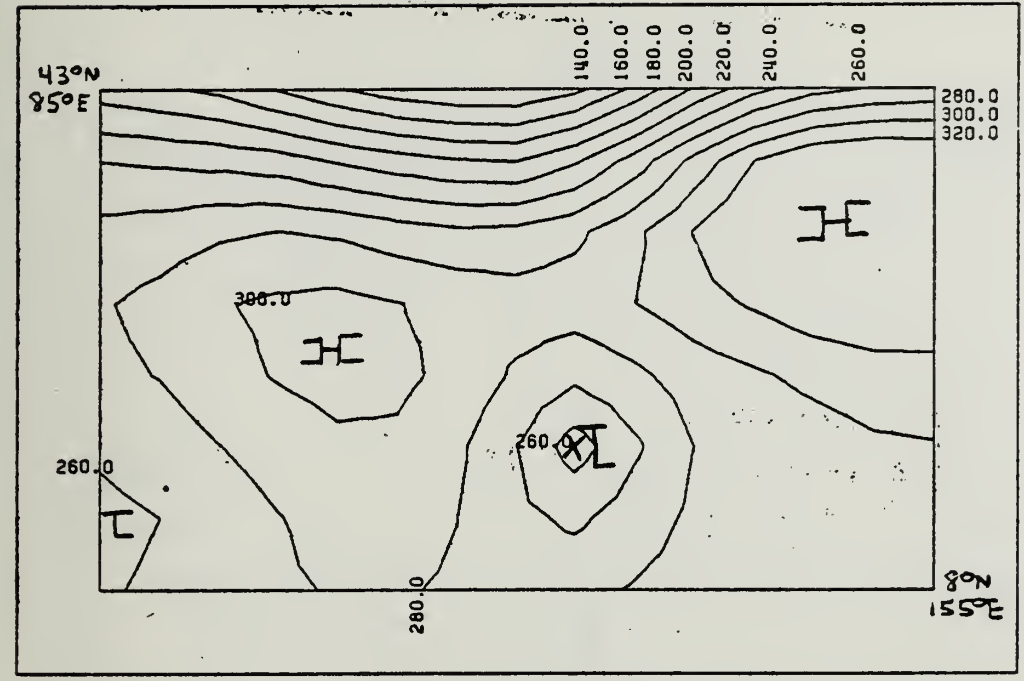


Fig. 4-18. Similar to Fig. 4-13, except first ten eigenvectors are used in reconstruction.

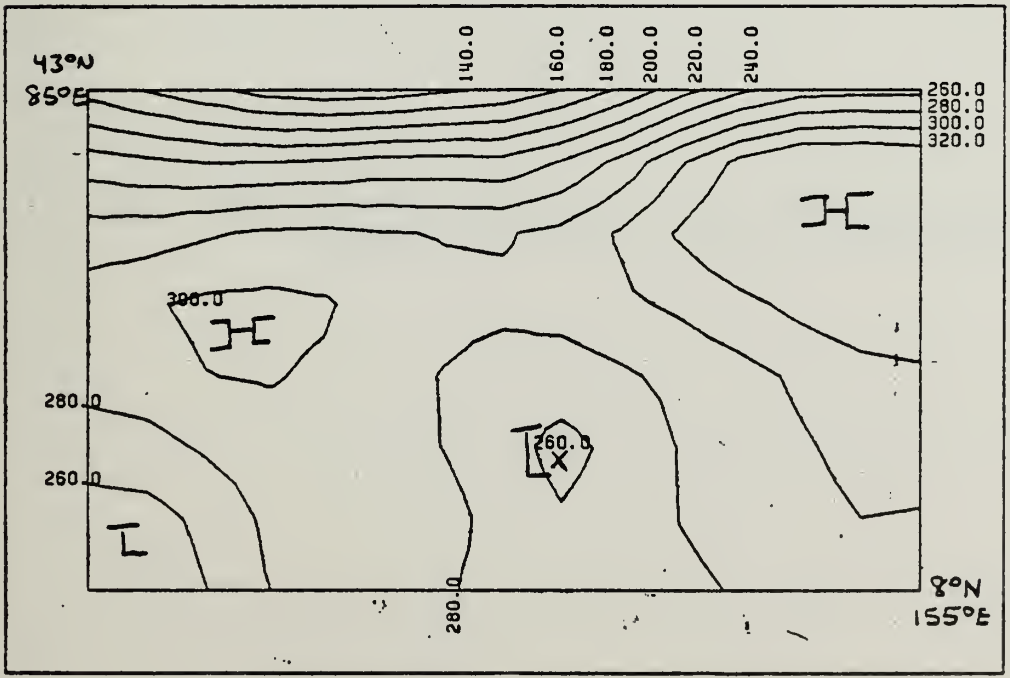


Fig. 4-19. Similar to Fig. 4-13, except first twenty eigenvectors are used in reconstruction.

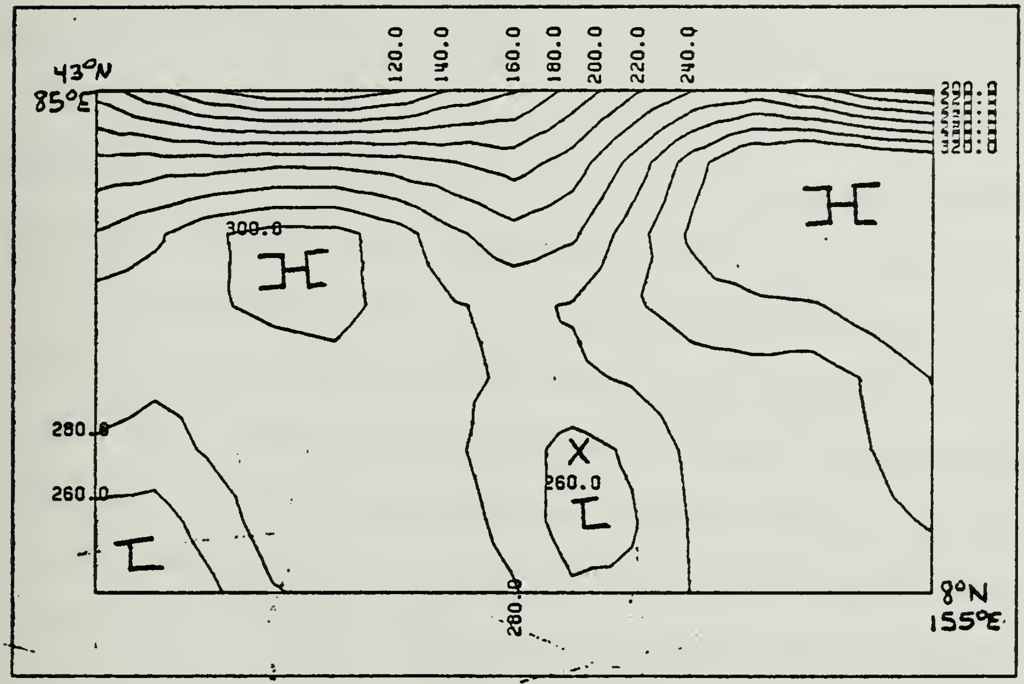


Fig. 4-20. Similar to Fig. 4-13, except first forty eigenvectors are used in reconstruction.

in agreement with the reconstruction shown above with the exception that the fourth instead of the fifth eigenvector seems to have a larger impact on the reconstruction.

Because inclusion of the eigenvectors 1, 3, 4, 5 and 9 seemed to have the greatest impact in the reconstruction, the orthogonal coefficients associated with these eigenvectors should have larger magnitudes than the other coefficients for this case. The values of the first ten coefficients are shown in Table 4-5. The coefficients associated with eigenvectors 1 and 9 are larger than the other coefficients. Although the value of coefficient 5 is the third largest value, it is the same magnitude as the coefficients associated with the second and third eigenvectors. This is explained in that eigenvector 2 tends to re-enforce the pattern of the first vector, while the third eigenvector enforces the joint pattern of one and two. The coefficient associated with the fourth eigenvector is small for this case, indicating that this pattern really had little effect on the reconstruction.

TABLE 4-4

Correlation coefficient of the reconstructed field, using the number of modes indicated, with the actual field being reconstructed (see text).

NUMBER OF MODES USED	1	2	3	4	5	6	7	8	9	10
CORRELATION	.618	.583	.663	.737	.752	.757	.728	.734	.885	.867

NUMBER OF MODES USED	15	20	25	30	40	50	60	120
CORRELATION	.852	.894	.936	.974	.994	.993	.994	1.000

TABLE 4-5

Values for the first 10 orthogonal coefficients for the case of 27 August 1967. (See text for details).

Coefficient	1	2	3	4	5	6	7	8	9	10
Value	5.94	1.50	-1.70	-.82	-1.85	-1.03	-.75	.26	2.56	-.38

These ten orthogonal coefficients define the pattern, and will be used shortly as predictors in regression equations for forecasting tropical cyclone motion. The hypothesis is that the forcing of typhoon motion may be determined from the various eigenvector patterns. As a preliminary test of this hypothesis, the zonal and meridional components of the typhoon motion (in nautical miles for various times) are correlated with the orthogonal coefficients associated with the eigenvectors (obtained from base time field). The correlations are calculated on 12-hour increments for the 12- to 84-hour displacement using the Pearson product moment (Dixon and Brown, 1979). Because the motion is defined to be positive to the north and to the west, a positive correlation means increased north or west forcing, relative to the mean displacement at a given time, with an increase in the value (not magnitude) of the coefficients. This holds for both the positive and negative (inverse) coefficients in that increases in value for a negative coefficient (decrease in magnitude) decreases the south or east forcing, or equivalently increases the north or west forcing. Each coefficient contributes to the total forcing, and the total movement is a summation of the forcing in all directions by all

eigenvectors. Correlations are obtained for a dependent set of 454 cases (or fewer for longer time intervals). Assuming the motion and orthogonal coefficients are both distributed normally, Chatfield (1980) shows the distribution of correlation coefficients for uncorrelated variables is distributed N(0,1/N). This means that any correlation of less than about .09 is not significant (at the 95% level). Tables 4-6 and 4-7 give the correlations for zonal and meridional motion, respectively.

Most of the correlations agree nicely with the instantaneous forcing of the eigenvectors inferred from Figs. 4-2 to 4-11, although there are surprises. Perhaps the largest surprise is the shift in meridional forcing in eigenvector 1 as the time interval increases. For times less than 36 hours, the forcing is the anticipated south forcing. The forcing at 48 and 60 hours is not significant, indicating the strength of this pattern at this time level gives little information on resultant 48- and 60-hour meridional motion. Between 72 and 84 hours, the forcing of this eigenvector actually becomes signficiantly northward from the mean 72 to 84 hour meridional displacement. A possible explanation for this phenomenon is that this pattern identifies recurving storms. During the short term, the forcing is to the south, but even more strongly to the west. The storm then crosses the mean meridional displacement location after 48 to 60 hours, still well to the west of the initial longitude. This is not to say the storm actually moves north of the initial latitude, only that the storm moves

Table 4 - 6

Pearson product moment (correlation) between the orthogonal coefficient associated with the given eigenvector and the zonal motion at 12 hour increments. A positive correlation implies west forcing. Also included is the instantaneous motion anticipated from the form of the eigenvectors in Figs 4-2 to 4-11.

MODE	ANTICIPATED FORCING	TIME INTERVAL						
		12	24	36	48	60	72	84
1	WEST	+.506	+.530	+.553	+.477	+.495	+.358	+.341
2	WEST	-.072	-.061	-.059	-.051	-.061	-.092	-.079
3	EAST	-.109	-.103	-.139	-.074	-.049	-.009	+.001
4	EAST	-.439	-.412	-.355	-.373	-.371	-.361	-.340
5	WEST	+.301	+.274	+.283	+.252	-.221	+.284	+.291
6	WEST	+.101	+.084	+.039	-.043	-.037	-.090	-.084
7	WEST	-.087	-.079	-.093	-.077	-.098	-.058	-.014
8	EAST	-.293	-.253	-.265	-.208	-.205	-.240	-.268
9	LITTLE	-.129	-.095	-.045	-.151	-.132	-.125	-.118
10	LITTLE	-.018	+.019	+.028	+.031	+.027	+.093	+.073

TABLE 4 - 7

Similar to Table 4 - 6, except for meridional motion and positive correlation implies northward forcing.

MODE	ANTICIPATED FORCING	TIME INTERVAL						
		12	24	36	48	60	72	84
1	SOUTH	-.199	-.211	-.242	+.017	+.056	+.194	+.312
2	SOUTH	-.213	-.184	-.175	-.175	-.158	-.205	-.164
3	NORTH	+.362	+.359	+.339	+.262	+.214	+.178	+.061
4	SOUTH	-.183	-.176	-.141	-.111	-.080	-.040	-.012
5	NORTH	+.075	+.034	+.017	+.009	-.005	+.037	-.047
6	LITTLE	-.158	-.163	-.136	-.068	-.112	-.102	-.122
7	NORTH	+.227	+.224	+.202	+.254	+.223	+.195	+.086
8	SOUTH	+.084	+.084	+.071	+.021	+.040	-.054	-.003
9	LITTLE	-.047	-.050	-.007	+.155	+.176	+.210	+.194
10	SOUTH	-.141	-.176	-.207	-.262	-.200	-.143	-.193

north of the expected latitudinal position at around 48 hours, and then remains north of the expected position. The westward forcing throughout the entire period is not inconsistent with recurvature, due to the large initial westward displacement. By the 72 hour time, the storm is north and west of the mean track displacement at that time, due only to coefficient 1 forcing. The storm displacement from the base time location is shown in Fig. 4-21 for all cases that have a 500mb coefficient 1 less than -9, while Fig. 4-22 is a graph of storm displacement for those storms with a coefficient 1 greater than +9. Recurvature is not seen immediately here, and more sophisticated statistical analysis techniques are required to verify the hypothesis presented above. Nevertheless, these two graphs show very nicely how the movement correlates with the coefficient value.

The other correlations shown in Tables 4-6 and 4-7 are consistent with the inferred instantaneous motion obtained from the eigenvectors. Eigenvectors 3 and 7 (along with 1) have the largest correlation (forcing) on the meridional motion. Eigenvector 1 has the greatest impact on the zonal forcing, with vectors 4, 5 and 8 also showing significant forcing. Surprisingly, eigenvectors 2 and 4 also correlated significantly with the meridional motion. From the results shown here, the anticipated forcing is in good agreement with the actual motion, and justifies use of the coefficients as predictors in regression equations for the storm motion.

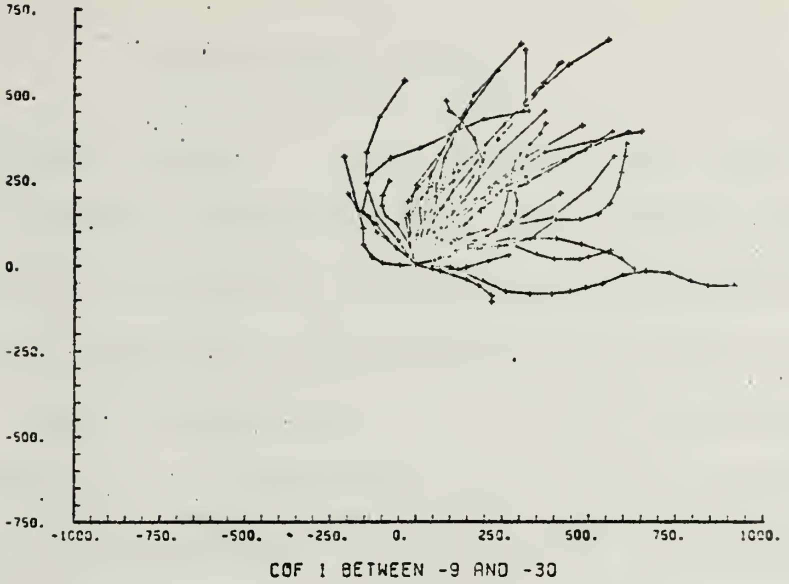


Fig. 4-21. Storm displacement from base time position, in nautical miles for all storms with 500mb coefficient 1 less than -9. 12-hour movement is indicated by a cross.

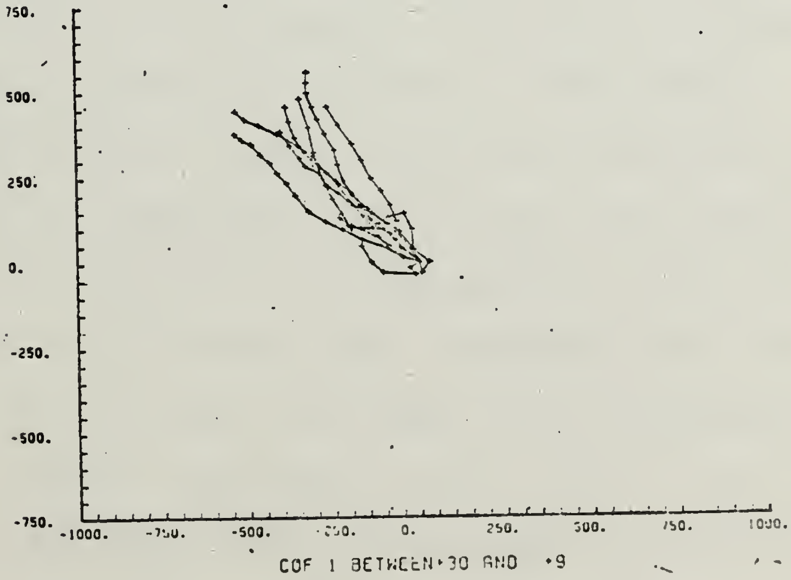


Fig. 4-22. Similar to Fig. 4-21 except these storms all have 500mb coefficient 1 greater than +9.

V. REGRESSION ANALYSIS

In the preceding chapter, it was demonstrated that the orthogonal coefficients associated with eigenvectors give qualitative insight to physical forcing mechanisms acting on tropical storms. Therefore, it is hypothesized that it is possible to use these coefficients to forecast quantitatively tropical storm motion. A regression approach is appropriate to investigate this hypothesis. Very briefly, regression analysis involves using a linear combination of known quantities (predictors) to estimate the value of an unknown quantity (predictand). Dixon and Brown (1979) give a concise summary of regression analysis, while Neter and Wasserman (1974) provide theoretical background of the technique. In the initial portion of this chapter, the model is developed, with model results appearing at the end of the chapter.

It was decided that of the 504 total data cases available, 50 would be used as independent cases to test the resultant equations. Use of 50 cases for the independent data set file is arbitrary, but still gives a large dependent data set. In the initial set of 504 cases, 185 cases had both complete past histories (warning positions 36 hours prior to the base time) and best track positions that extended to 84 hours beyond the base time. Of these 185 cases, it was decided to hold 35 cases to comprise part of the independent set, leaving 150 cases with full history in the dependent set. The remaining

15 independent cases were selected from the remaining cases without complete history. All cases in independent data set were selected randomly within their respective history subsets. This process left 454 potential cases over which the regression equations were formed. The fifty independent cases are shown in Table 5-1. It will be shown shortly that the actual number of cases used to derive the regression equations is less than 454, due to the specifications of the predictors.

Predictands for this study are the 12- to 84-h zonal and meridional displacements of the storms in 12-hour increments. These distances are determined from the base time JTWC warning position to the JTWC best-track position at the predictand time. Positive motion is defined to the north and to the west, since the majority of the displacements are to the north and west. As there are 14 predictands, 14 regression equations are required for each of the three pressure levels for which synoptic data are available. Because the basic data are only available at 12-hour intervals, and the analyzed maps are delayed several hours, the forecast time must be carefully distinguished from the guidance time. A 12-h forecast based on 0000GMT data is the forecast position valid at 1200GMT, whereas a 12-h guidance based on the 0000GMT data would be issued several hours after 0000GMT and would be valid 12 hours after issuance. It is estimated that four hours would be needed to prepare and issue the forecast. Hence, a forecast issued based on 0000GMT data could only be used in preparing

TABLE 5 - 1

The independent storms; their dates of occurrence, position and intensity, and their past warning and future best track history.

	NAME	YEAR	MONTH/DATE	TIME	LAT	LON	MAX WIND	HOURS PRIOR	FUTURE POSITION
1	THERESE	1967	3 / 18	000GMT	10.7N	139.9E	40	36	84
2	VIOLET	1967	4 / 3	1200GMT	10.9N	138.3E	65	36	84
3	GEORGIA	1967	7 / 31	000GMT	22.0N	136.7E	35	36	84
4	GEORGIA	1967	8 / 6	000GMT	35.9N	150.1E	50	36	24
5	OPAL	1967	9 / 10	000GMT	26.6N	140.4E	85	36	84
6	RUTH	1967	9 / 9	000GMT	27.1N	162.8E	55	36	84
7	DINAH	1967	10 / 19	000GMT	10.6N	138.7E	60	36	84
8	GILDA	1967	11 / 11	000GMT	10.6N	152.9E	75	36	84
9	JEAN	1968	4 / 9	1200GMT	10.6N	150.6E	85	36	84
10	KIM	1968	6 / 2	000GMT	17.5N	132.5E	85	36	84
11	WENDY	1968	8 / 31	1200GMT	20.5N	141.9E	130	36	84
12	AGNES	1968	8 / 31	1200GMT	16.1N	155.9E	75	36	84
13	AGNES	1968	9 / 8	000GMT	23.4N	137.2E	60	36	48
14	DELLA	1968	9 / 21	000GMT	19.9N	128.1E	105	36	84
15	CARMEN	1968	9 / 18	000GMT	18.2N	147.2E	70	36	84
16	JUDY	1968	10 / 26	1200GMT	11.0N	147.8E	100	36	84
17	JUDY	1968	10 / 29	1200GMT	16.6N	135.6E	105	36	72
18	HELEN	1969	10 / 11	000GMT	23.7N	141.7E	95	36	36
19	IDA	1969	10 / 18	000GMT	18.8N	145.6E	90	36	84
20	GRACE	1969	10 / 1	000GMT	26.9N	166.6E	70	12	48
21	GRACE	1969	10 / 2	1200GMT	24.7N	162.8E	70	36	84
22	BILLIE	1970	8 / 27	1200GMT	27.0N	131.3E	110	36	84
23	JCAN	1970	10 / 14	1200GMT	14.4N	117.5E	85	36	84
24	PATSY	1970	11 / 20	1200GMT	15.7N	114.7E	60	36	24
25	MARGE	1970	11 / 3	000GMT	14.7N	116.9E	55	24	60
26	VERA	1971	4 / 15	1200GMT	18.2N	125.6E	85	36	48
27	WANDA	1971	4 / 29	1200GMT	11.7N	112.1E	40	36	84
28	BABE	1971	5 / 5	000GMT	19.2N	119.3E	45	36	48
29	LUCY	1971	7 / 19	1200GMT	18.7N	124.7E	125	36	60
30	TRIX	1971	8 / 22	1200GMT	25.7N	151.0E	75	36	84
31	TRIX	1971	8 / 25	1200GMT	25.2N	142.9E	85	36	84
32	VIRGINIA	1971	9 / 4	1200GMT	22.2N	136.9E	60	36	60
33	WENDY	1971	9 / 9	1200GMT	24.1N	158.3E	105	36	60
34	EMMA	1974	6 / 15	000GMT	15.7N	127.0E	40	36	48
35	PCLLY	1974	8 / 28	000GMT	19.8N	143.5E	75	36	84
36	AGNES	1974	9 / 26	1200GMT	24.9N	151.9E	50	36	84
37	ELAINE	1974	10 / 27	000GMT	16.9N	127.1E	85	36	84
38	GLORIA	1974	11 / 5	000GMT	15.6N	131.2E	85	36	84
39	IRMA	1974	11 / 24	1200GMT	14.6N	134.3E	70	36	84
40	LOLA	1975	1 / 25	1200GMT	12.3N	117.0E	45	36	48
41	RITA	1975	8 / 20	000GMT	26.5N	128.8E	45	36	84
42	GRACE	1975	10 / 29	1200GMT	17.9N	128.8E	30	36	84
43	RITA	1972	7 / 15	000GMT	21.1N	135.6E	80	36	84
44	RITA	1972	7 / 16	1200GMT	21.8N	134.8E	65	36	84
45	TESS	1972	7 / 17	000GMT	21.1N	151.7E	110	36	84
46	ALICE	1972	8 / 6	000GMT	30.2N	144.2E	75	36	36
47	CLGA	1976	5 / 16	1200GMT	12.3N	129.8E	45	36	84
48	SALLY	1976	6 / 28	000GMT	19.4N	132.0E	100	36	84
49	THERESE	1976	7 / 15	1200GMT	22.4N	136.9E	120	36	84
50	BILLIE	1973	7 / 15	000GMT	20.9N	125.3E	115	36	84

the 0400GMT guidance. A 12-h guidance will then be valid at 1600GMT. To insure that an estimate of the position during the next 72 hours is always available, forecasts are made to 84-h after the base time. All subsequent references to times will be for forecast rather than guidance timing.

The potential predictors are identical for all of the 14 regression equations, with the exception of any predictors that are a function of atmospheric level. Predictors are sought to assess quantitatively the effect of three different features on storm movement: external (to the storm) physical forcing, previous movement of the storm, and storm intensity. Synoptic (and sub-synoptic) external forcing on the storm is thought to play a large role on storm movement (Brown, 1981 and others). To incorporate the forcing quantitatively, the orthogonal coefficients associated with the 10 retained eigenvectors for a particular data case are selected as potential predictors. One of the primary objectives in this study is to determine how well these EOF's represent large scale features.

If the storm is to be forecast properly, prior motion must also be accounted for (Peterson, 1980). It is necessary to know toward which direction the storm is moving to determine what portion of the external forcing will be affecting the storm. To do this, twelve additional variables representing past zonal and meridional displacements are added to the set of potential predictors. All of the prior storm displacements are based on warning positions to simulate operational

conditions. The six variables for zonal motion are the prior 12, 24 and 36 hour zonal displacements of the storm, the zonal displacements from 12 hours to 24 and 36 hours prior, and finally the zonal displacements from 24 to 36 hours prior to the base time. The time frames for the meridional displacements are identical.

Storm intensity is the third storm characteristic sought to assess quantitatively. The most preferable form of this data would be a meso- or microscale analysis of the winds around the storm. Since this is not available, the JTWC warning maximum winds are used to indicate intensity. The intensity data are available for the base time, and at 12, 24 and 36 hours prior to base time. Therefore, the complete set of potential predictors includes four predictors for intensity, 12 for past movement and 10 for the physical forcing. Table 5-2 is a listing of the 26 potential predictors, along with the names used to identify each predictor in this study. For a data case to be used in the formulation of the regression equations, a complete set of potential predictors and the proper predictand had to be available. This decreased the number of cases available for computation of the regression equations. Actual valid case numbers are presented with the results of the regression. Since the number of potential predictors is initially large, the resultant equations need to be examined carefully to determine if any of these predictors may be excluded with little information loss. It is

TABLE 5 - 2

Potential predictors used to develop the regression equations. The first ten predictors are different for each of the three pressure levels.

POTENTIAL PREDICTOR VARIABLE NUMBER	NAME	DESCRIPTION
1	cof1	The orthogonal coefficient associated with eigenvector 1.
2	cof2	The orthogonal coefficient associated with eigenvector 2.
3	cof3	The orthogonal coefficient associated with eigenvector 3.
4	cof4	The orthogonal coefficient associated with eigenvector 4.
5	cof5	The orthogonal coefficient associated with eigenvector 5.
6	cof6	The orthogonal coefficient associated with eigenvector 6.
7	cof7	The orthogonal coefficient associated with eigenvector 7.
8	cof8	The orthogonal coefficient associated with eigenvector 8.
9	cof9	The orthogonal coefficient associated with eigenvector 9.
10	cof10	The orthogonal coefficient associated with eigenvector 10.
11	plat1	Storm latitude movement for 12 hours before base time.
12	plat2	Storm latitude movement for 24 hours before base time.
13	plat3	Storm latitude movement for 36 hours before base time.
14	plat4	Storm latitude movement from 24 to 12 hours before base time.
15	plat5	Storm latitude movement from 36 to 12 hours before base time.
16	plat6	Storm latitude movement from 36 to 24 hours before base time.
17	plon1	Storm longitude movement for 12 hours before base time.
18	plon2	Storm longitude movement for 24 hours before base time.
19	plon3	Storm longitude movement for 36 hours before base time.
20	plon4	Storm longitude movement from 24 to 12 hours before base time.
21	plon5	Storm longitude movement from 36 to 12 hours before base time.
22	plon6	Storm longitude movement from 36 to 24 hours before base time.
23	amw0	Storm warning maximum wind at forecast base time.
24	amw1	Storm warning maximum wind 12 hours prior to base time.
25	amw2	Storm warning maximum wind 24 hours prior to base time.
26	amw3	Storm warning maximum wind 36 hours prior to base time.

desirable to have as few potential predictors as possible. Therefore, if it is determined that any of the potential predictors add little to the equations, they should be dropped from the developmental set, and the equations should be rederived over the smaller set of predictors.

The next decision is how to use the predictors to create the equations. Two primary possibilities exist: all possible predictors or stepwise regression. All possible predictor regressions use all predictors at once to form the regression equations. In this study, all 26 predictors would be used to formulate the equations. A stepwise regression creates the regression equations by adding (or deleting) one predictor per step. At each step, the single predictor that is most highly correlated with any residual error from the previous step is added to the predictors used, and the equations (and residuals) recomputed. This process continues until no additional predictors meet a pre-assigned significance tolerance level. Dixon and Brown (1979) give further details of the procedure. Typically, not all potential predictors are used.

A stepwise screening procedure is used here for two fundamental reasons. First, a stepwise procedure extracts maximum information out of minimum variables, and variables that add little information are not used. Second, and more importantly, Neter and Wasserman (1974) show that if two or more potential predictors are highly correlated, retention of both may have a deleterious effect on interpretation of the equations.

The problem is called multicollinearity. Statistically, the effect is to have little additional reduction in the total explained variance, while decreasing the degrees of freedom in the equation. Since at least some of the potential predictors are highly correlated, multicollinearity could be a problem. By using a stepwise regression approach, the problem is circumvented. Whenever a stepwise regression scheme is used, a decision on how many predictors are to be used needs to be made. Two possible approaches are to use a predetermined number of predictors, so that the number of terms in each final equation are identical, or to use all terms that meet a predetermined significance tolerance level. For this study, all predictors that significantly reduce the variance are included in the equations, so that the number of terms in the various equations differs. A tolerance level (F-ratio) of 4.0 is used for this study (Dixon and Brown, 1979).

Finally, the form of the equations, either linear or polynomial, must be decided. The simplest type of polynomial regression involves using all first-order predictors, and nonlinear combinations of the first-order predictors in the model. For instance, if there are 10 initially defined potential predictors, then the set of predictors used in polynomial regression include all 10 first order terms, all 10 second order (squared) predictors, plus the 45 nonlinear products of all potential predictors. The use of polynomial regression may occasionally be of aid in fitting the predictors to the

predictands when nonlinear cause and effect is anticipated. Neumann and Leftwich (1977) use a second order polynomial regression to forecast typhoon movement, although their predictors do not include synoptic forcing explicitly. With 26 potential predictors, as in this study, the number of polynomial predictors becomes unwieldy. A further justification for not using polynomial regression is that the predictands give no evidence of interacting nonlinearly with the predictors.

In summary, 14 linear regression equations are to be formulated for each atmospheric pressure level, with predictands being 12- through 84-h zonal and meridional displacements (in nautical miles) in 12-hour increments. Predictors will be selected stepwise from a set of 26 potential predictors over 454 (or fewer) dependent data cases. 50 cases have been held back to test the equations.

The regression equations are calculated using the University of California BMDP computer routine linear stepwise regression (Dixon and Brown, 1979). Before presenting the equations, their ability to explain variation in the predictand is examined by use of R^2 statistic. This quantity may be interpreted as the percent explained variance in the predictand by the regression equation (using the dependent data cases). The R^2 value for each regression equation is shown in Table 5-3.

Several properties are immediately seen from the R^2 values. First, the zonal equations appear to explain a greater portion

TABLE 5 - 3

Sample size and R^2 statistic for each zonal and meridional regression equation by forecast time and atmospheric level.

	FORECAST INTERVAL (HR)						
	12	24	36	48	60	72	84
NUMBER OF DEPENDENT DATA CASES	351	351	329	256	233	163	150
ZONAL EQUATIONS							
500mb	.794	.725	.685	.613	.568	.556	.444
700mb	.791	.719	.680	.600	.558	.550	.310
850mb	.784	.712	.651	.571	.519	.535	.384
MERIDIONAL EQUATIONS							
500mb	.522	.476	.404	.354	.255	.315	.208
700mb	.540	.486	.419	.347	.285	.252	.184
850mb	.502	.463	.365	.323	.255	.259	.103

of the total (zonal) movement variation than do the meridional equations. Over 75% of the total (zonal) variation in the 12-h movement is explained by the equations at each of the three atmospheric levels. The maximum meridional variation explained (54%) is for the 12-h movement using 700mb EOF coefficients. Matching forecast times and levels (excluding the 84 hour forecast from the 700mb equations), the zonal R^2 is always at least .24 greater than the meridional R^2 for the same time period and level. The increased ability of the zonal equations is expected because there is greater variation in the zonal movement than the meridional movement. The means and standard deviations of the zonal and meridional displacements at the various forecast times are shown in Table 5-4.

TABLE 5-4

Means and standard deviations of the predictands (in nautical miles) for the dependent sample. See text for details.

	FORECAST TIME (HOURS)						
	12	24	36	48	60	72	84
Meridional displacement							
mean	56	119	181	223	282	316	353
standard deviation	(50)	(100)	(150)	(165)	(221)	(230)	(267)
Zonal displacement							
mean	51	93	129	195	225	307	372
standard deviation	(81)	(176)	(258)	(309)	(376)	(396)	(449)

The mean movement for both directions is roughly the same magnitude, and indicates an average track toward the northwest. A more significant difference in the motion is seen in the standard deviations, which are larger for the zonal motion than for the meridional motion. As both the zonal and meridional components contribute approximately the same error magnitude in the regression equations, the R^2 for the zonal motion will be significantly greater since there is more variance to be explained.

The second property seen immediately in the R^2 values in Table 5-3 is that they decrease rapidly in time for each pressure level. For the 500mb equations, a general rule of thumb is that the R^2 decreases by a value of .05 per 12 hour increment. It is further seen (Table 5-4) that the standard deviation of displacement increases every 12 hours, heightening the significance of the decrease of the R^2 in time. Simply stated, the equations predict movement well in the short term, but the errors grow rapidly with increasing time.

The final property seen in the R^2 values is that the accuracy of the equations is not a strong function of the atmospheric level in the dependent sample case. The 500mb R^2 values are generally larger than at the other two levels, although these differences are not significant. A Student's t-test, assuming non-identical variacnes in the population, was conducted with the null hypothesis that there is no significant difference in the R^2 values at the various levels.

In no case was the test statistic significant at even the alpha equal .75 level. Therefore, the null hypothesis is accepted that over the dependent sample there is no difference in performance of the equations at the different atmospheric levels.

Tables 5-5 and 5-6 present the regression coefficients of the 500mb equations by direction of movement. For example, the 500mb meridional regression coefficients for all seven forecast times are given in Table 5-5. The first value given is the intercept. The final regression equation prediction of displacement is obtained by summing over the product of all non-zero regression coefficients and the variable associated with the coefficient. None of the 500mb equations use more than 10 predictors. In seven of the 28 equations, six or fewer predictors are used. Therefore, these equations are very simple to use. A past movement variable was always the first variable selected in the stepwise procedure, so persistence does play a role in the predicted movement. The predictions are not simply persistence forecasts, however, since in general four or five EOF coefficient predictors are chosen in each equation. Therefore, forcing also plays a crucial role in the storm movement. Finally, maximum wind predictors are of little consequence in the final equations, indicating little impact on the 12-h (or greater) time scale storm motion (excluding short term trochoidal path oscillation). The resultant equations for the 700 and 850mb data are shown in Appendix B. It is also noted that of the potential

TABLE 5 - 5

Regression coefficients for the seven meridional equations using 500mb EOF's. A value of .0 indicates the predictor was not selected in the stepwise selection procedure.

	FORECAST VALID FOR BASE TIME PLUS HOURS						
	12	24	36	48	60	72	84
Intercept	38.334	70.789	123.114	117.334	214.492	182.498	297.612
Cof1	.0	.0	-2.886	.0	.0	9.738	20.376
Cof2	-2.234	-4.435	-6.148	-5.649	-6.633	-11.960	.0
Cof3	3.848	8.781	13.491	12.635	12.677	11.923	.0
Cof4	-2.641	-5.799	-8.169	.0	.0	.0	.0
Cof5	.0	.0	.0	.0	.0	.0	.0
Cof6	-2.535	-5.279	-7.191	.0	-14.631	.0	.0
Cof7	3.182	6.502	10.390	17.320	26.948	20.428	.0
Cof8	.0	.0	.0	.0	.0	.0	.0
Cof9	.0	.0	.0	12.293	18.113	24.462	28.487
Cof10	-2.618	-8.975	-18.292	-16.197	.0	.0	.0
Plat1	.0	0.634	0.652	1.068	1.405	1.247	.0
Plat2	0.358	.0	.0	.0	.0	.0	0.656
Plat3	.0	.0	.0	.0	.0	.0	.0
Plat4	-0.286	.0	.0	.0	.0	.0	.0
Plat5	.0	.0	.0	.0	.0	.0	.0
Plat6	.0	.0	.0	.0	.0	.0	.0
Plon1	0.246	0.502	0.257	.0	.0	.0	.0
Plon2	-0.088	-0.158	.0	.0	.0	.0	.0
Plon3	.0	.0	.0	.0	.0	0.197	.0
Plon4	.0	.0	.0	.0	.0	.0	.0
Plon5	.0	.0	.0	.0	.0	.0	0.332
Plon6	.0	.0	.0	.0	.0	.0	.0
Amw0	.0	.0	.0	.0	.0	.0	.0
Amw1	.0	0.319	0.518	0.685	.0	1.351	.0
Amw2	.0	.0	.0	.0	.0	.0	.0
Amw3	.0	.0	.0	.0	.0	.0	.0

TABLE 5 - 6

Regression coefficients for the seven zonal equations using 500mb EOF's. A value of .0 indicates the predictor was not selected in the stepwise selection procedure.

	FORECAST VALID FOR BASE TIME PLUS HOURS						
	12	24	36	48	60	72	84
Intercept	16.027	37.064	36.833	105.903	216.515	168.503	286.962
Cof1	2.678	6.664	13.668	18.466	26.153	19.369	27.915
Cof2	.0	.0	.0	.0	.0	.0	.0
Cof3	.0	.0	-6.104	.0	.0	.0	.0
Cof4	-3.635	-7.783	-11.698	-21.928	-32.626	-48.063	-51.194
Cof5	4.239	8.460	12.385	11.074	.0	24.253	32.448
Cof6	.0	.0	.0	.0	.0	.0	.0
Cof7	.0	.0	.0	.0	.0	.0	.0
Cof8	-7.484	-12.328	-22.332	-22.361	-26.982	-41.319	-58.377
Cof9	.0	.0	.0	.0	.0	.0	.0
Cof10	.0	.0	13.350	.0	.0	34.037	.0
Plat1	.0	.0	.0	-0.758	-1.058	.0	.0
Plat2	.0	.0	.0	.0	.0	-0.660	-0.836
Plat3	.0	.0	.0	.0	.0	.0	.0
Plat4	.0	.0	.0	.0	.0	.0	.0
Plat5	.0	.0	.0	.0	.0	.0	.0
Plat6	.0	-0.234	.0	.0	.0	.0	.0
Plon1	-0.626	-1.232	-1.593	-1.782	-1.919	-1.798	-1.542
Plon2	.0	.0	.0	.0	.0	.0	.0
Plon3	.0	.0	.0	.0	.0	.0	.0
Plon4	.0	.0	.0	.0	.0	.0	.0
Plon5	.0	.0	.0	.0	.0	.0	.0
Plon6	.0	.0	.0	.0	.0	.0	.0
Amw0	.0	.0	.0	.0	.0	2.179	.0
Amw1	.0	.0	.0	.0	.0	.0	.0
Amw2	.0	.0	.0	.0	-1.165	.0	.0
Amw3	.0	.0	.0	.0	.0	-2.113	.0

predictors, very little information would be lost by excluding all past displacement variables except for the 12-h period prior to base time. Additionally, of the intensity predictors, the most frequently selected was the 12 hour prior intensity. Therefore, it was decided to rederive the equations using only 13 potential predictors (the 10 coefficients at the given level, Platl, Plonl and Amwl). Results of the equations, in the form of R^2 statistics, derived on the smaller set are given in Appendix 3. The remainder of the results presented in this chapter refer to the equations derived using the complete set of all 26 potential predictors.

Results presented thus far have been drawn from the regression equations using the dependent data set. A true test of a regression equation comes through testing with independent data. This testing is critical in determination of accuracy of the model. The JTWC annual typhoon report publishes, in addition to best track and warning positions, the forecast errors for 24, 48 and 72 hour forecasts. The regression model was tested with the independent data and is compared to the official JTWC forecast error, which serves as a benchmark. Of the 50 independent cases, only 45 have JTWC official forecasts at 24 hours, 31 have official forecasts at 48 hours and only 17 at 72 hours. Admittedly, the sample size of the independent storms is quite small, but inferences on aptness of the model may still be

drawn. Both the complete set of results for the independent storms, and the homogeneous set where both JTWC and the regression model errors are available will be shown.

The overall performance (Table 5-7) of the regression equations on the entire set of 50 independent cases is first examined to determine if there is consistency in the forecasts (indicated by small standard deviations) and to determine in general how well the equations forecast the motion.

TABLE 5-7

Mean and standard deviation forecast vector error (nautical miles) of 24, 48 and 72 hours for the set of 50 independent storms.

	HOUR FORECAST		
	24	48	72
Sample size	50	43	36
500mb forecast error			
mean	88.4	176.4	277.4
standard deviation	62.5	113.5	167.4
700mb forecast error			
mean	110.1	189.3	318.7
standard deviation	91.3	120.5	178.7
850mb forecast error			
mean	114.9	205.4	358.0
standard deviation	105.8	146.1	219.2

The 500mb equations outperformed the other two equation sets by a wide margin, which is surprising. Similar differences between levels did not appear in the errors of the dependent

sample, given in Table 5-8. A possible explanation is that there is a greater variation in the synoptic forcing fields at 500mb. This allows the 500mb equations to be less susceptible to large forecast errors in cases where the predictors have extreme values. It turns out that with few exceptions, the 700mb errors are similar to the 500mb errors. Where the 700mb equations performed poorly, the results were much worse than the 500mb equations. Therefore, it appears that (at least over the independent cases) the 500mb equations have a smaller likelihood to give a large forecast error. This hypothesis needs to be tested more thoroughly as additional data becomes available.

TABLE 5-8

Mean and standard deviation forecast vector error (nautical miles) of 24, 48 and 72 hours for the set of 454 dependent storms.

	FORECAST INTERVAL		
	24	48	72
Sample size	351	255	164
500mb forecast error			
mean	91.5	203.3	298.7
standard deviation	72.7	113.7	152.4
700mb forecast error			
mean	92.6	210.6	293.7
standard deviation	71.9	115.8	121.5
850mb forecast error			
mean	95.2	210.7	383.4
standard deviation	71.6	121.5	232.2

The next step in examination of the independent data results is to compare the results of EOF regression forecasts to the official JTWC forecasts, for those cases that this is possible. The mean and standard deviation errors for these valid cases, and the benchmark JTWC official forecast error statistics are shown in Table 5-9. A superior 500mb scheme is again evident. More importantly, it is seen the standard deviation of error for the EOF regression scheme is less than for the JTWC official forecasts, which indicates the EOF regression scheme is less likely to have a large forecast error. The combination of small mean error and small standard deviation indicates the EOF scheme outperforms the JTWC official forecast. The 700 and 850mb equation forecasts were again poorer than the 500mb forecasts, and appear to be about equal to the JTWC forecasts.

Finally, the EOF regression scheme is compared to the JTWC official forecast on a case-by-case basis in Figs. 5-1 through 5-9. Any points lying above the straight line on the graphs represent cases in which the EOF scheme outperformed the JTWC official forecasts. The 850mb results (Figs. 5-3, 5-6 and 5-9) show little differences between the schemes. The 700mb equations (Figs. 5-2, 5-5 and 5-8) show, in general, a better forecast by the EOF scheme, as a bulk of the points lie above the no difference line. The overall comparison statistics appear to have been affected by a few large forecast errors, especially at 24 hours. This tendency

TABLE 5 - 9

Mean and standard deviation of forecast vector magnitude error (n. mi.) for the EOF regression scheme and the JTWC official forecast for the independent storms. Only those storms where both forecasts have valid errors are compared.

	HOUR FORECAST		
	24	48	72
NUMBER VALID COMPARISION CASES	45	31	17
JTWC official forecast forecast error			
mean	110.6	231.8	329.2
standard deviation	68.9	157.2	222.4
500mb forecast error			
mean	89.4	197.1	274.3
standard deviation	64.4	128.0	224.4
700mb forecast error			
mean	107.0	209.1	297.2
standard deviation	89.3	132.4	233.1
850mb forecast error			
mean	113.5	208.1	318.0
standard deviation	102.2	128.2	259.5

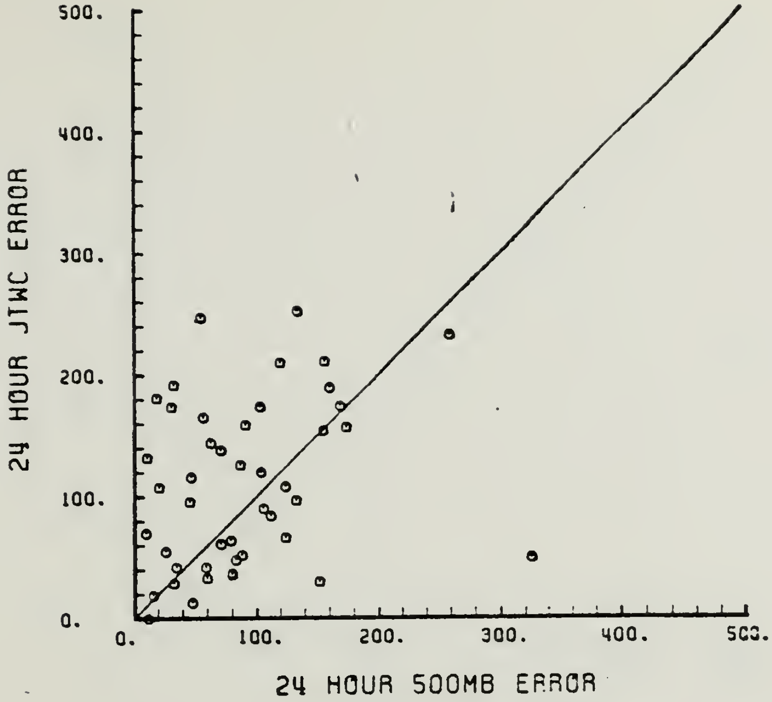


Fig. 5-1. Comparison of the forecast error for the independent data cases. Schemes compared are the 500mb EOF regression scheme versus the JTWC official forecast, for a 24 hour forecast. Units are in nautical miles.

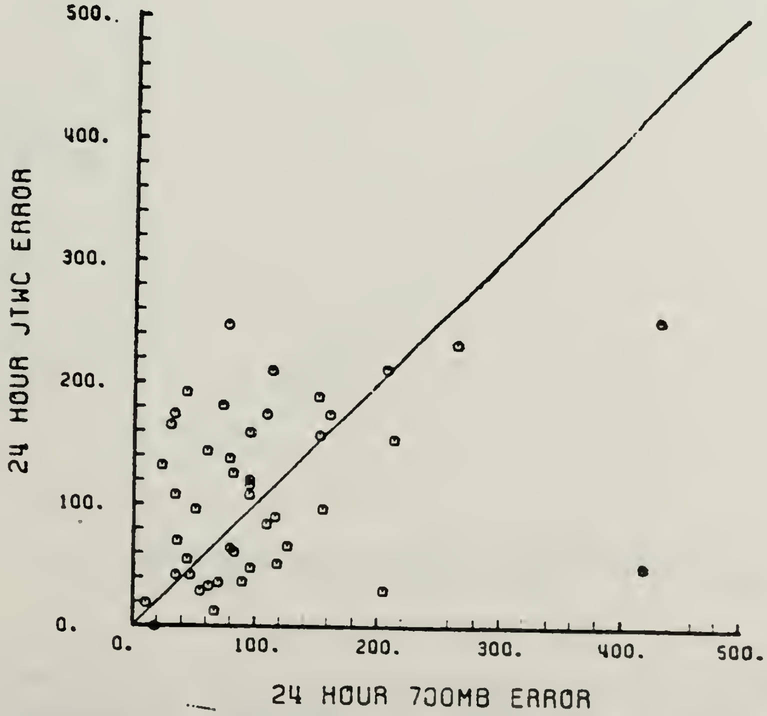


Fig. 5-2. Similar to Fig. 5-1, except the 700mb EOF regression forecast is compared to JTWC official forecast for a 24-hour forecast.

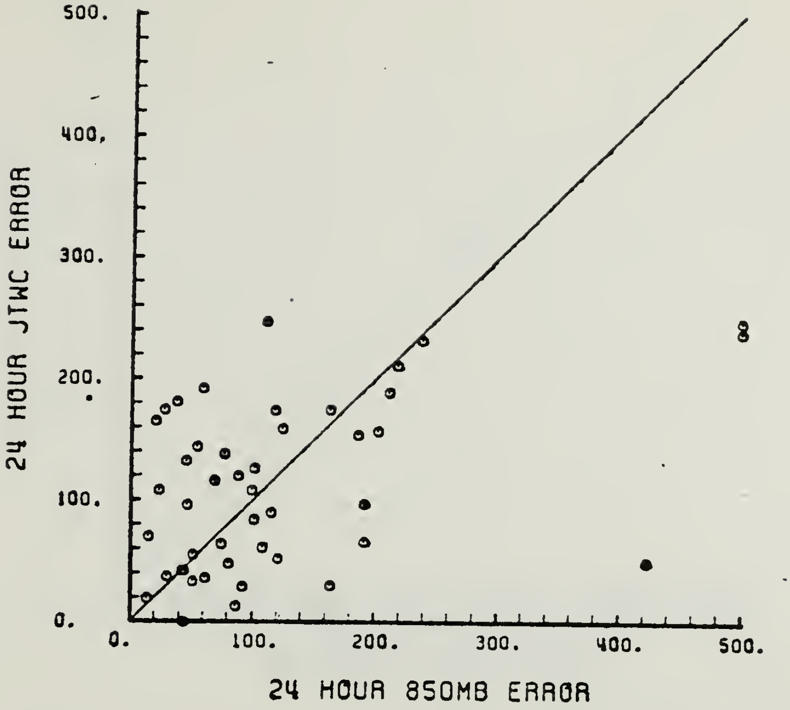


Fig. 5-3. Similar to Fig. 5-1, except the 850mb EOF regression forecast is compared to JTWC official forecast for a 24-hour forecast.

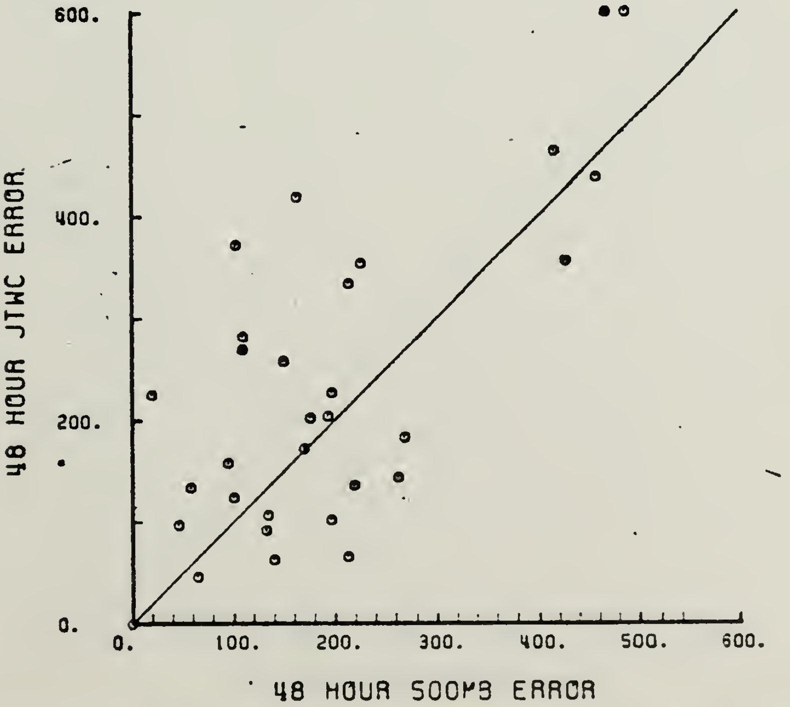


Fig. 5-4. Similar to Fig. 5-1, except the 500mb EOF regression forecast is compared to JTWC official forecast for a 48-hour forecast.

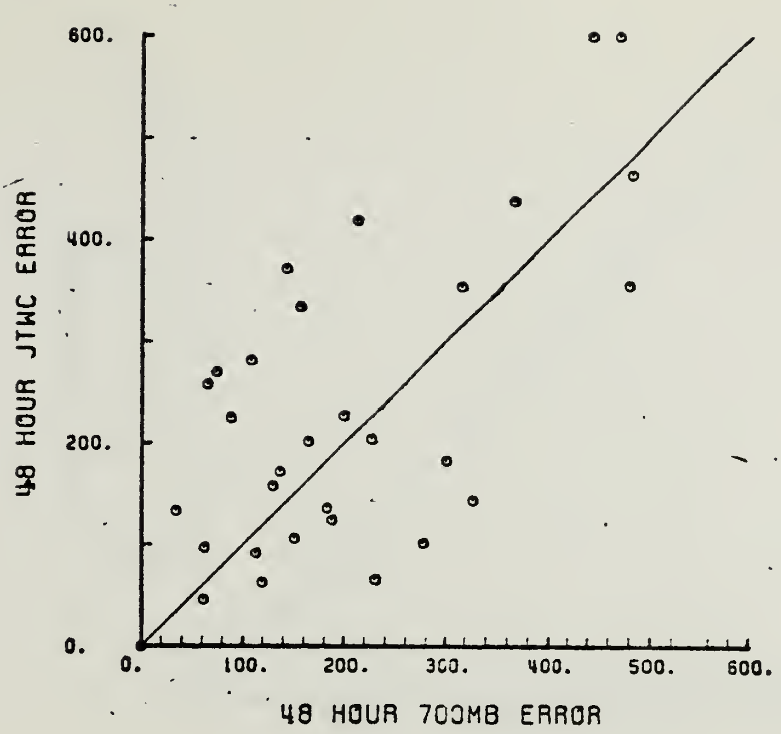


Fig. 5-5. Similar to Fig. 5-1, except the 700mb EOF regression forecast is compared to JTWC official forecast for a 48-hour forecast.

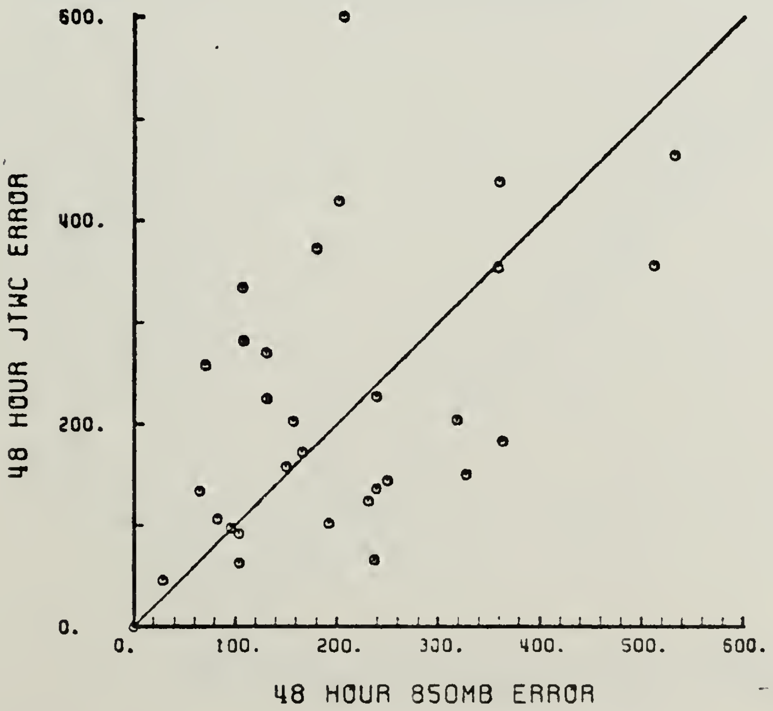


Fig. 5-6. Similar to Fig. 5-1, except the 850mb EOF regression forecast is compared to JTWC official forecast for a 48-hour forecast.

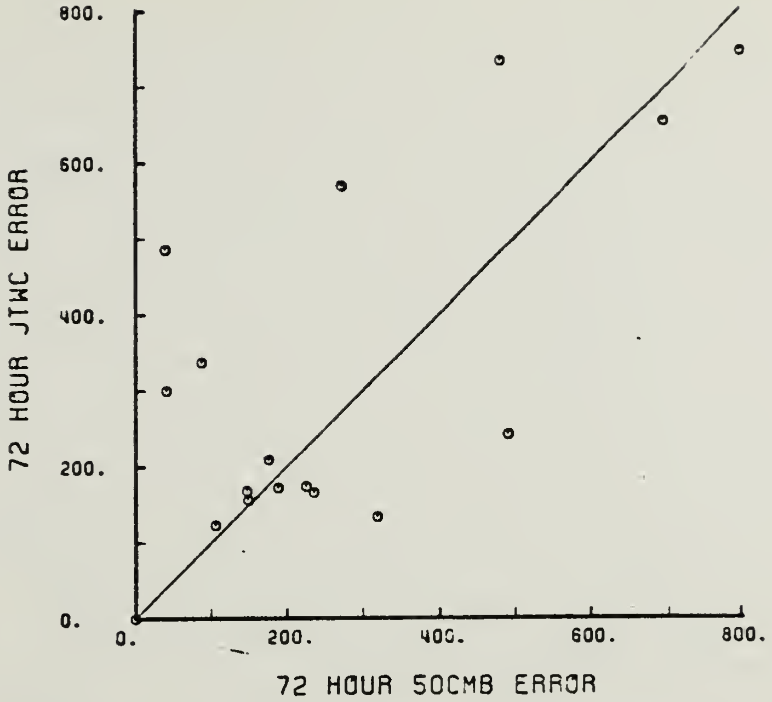


Fig. 5-7. Similar to Fig. 5-1, except the 500mb EOF regression forecast is compared to JTWC official forecast for a 72-hour forecast.

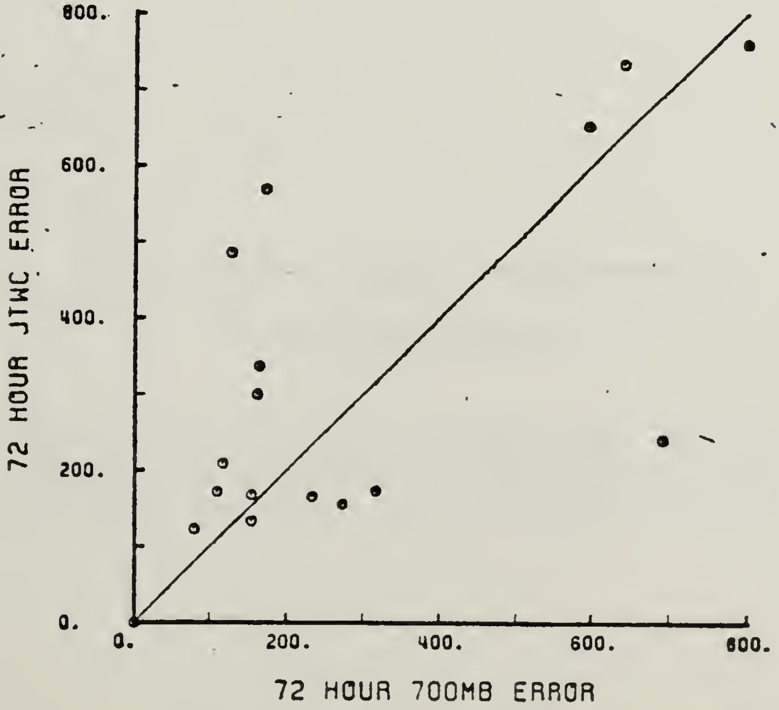


Fig. 5-8. Similar to Fig. 5-1, except the 700mb EOF regression forecast is compared to JTWC official forecast for a 72-hour forecast.

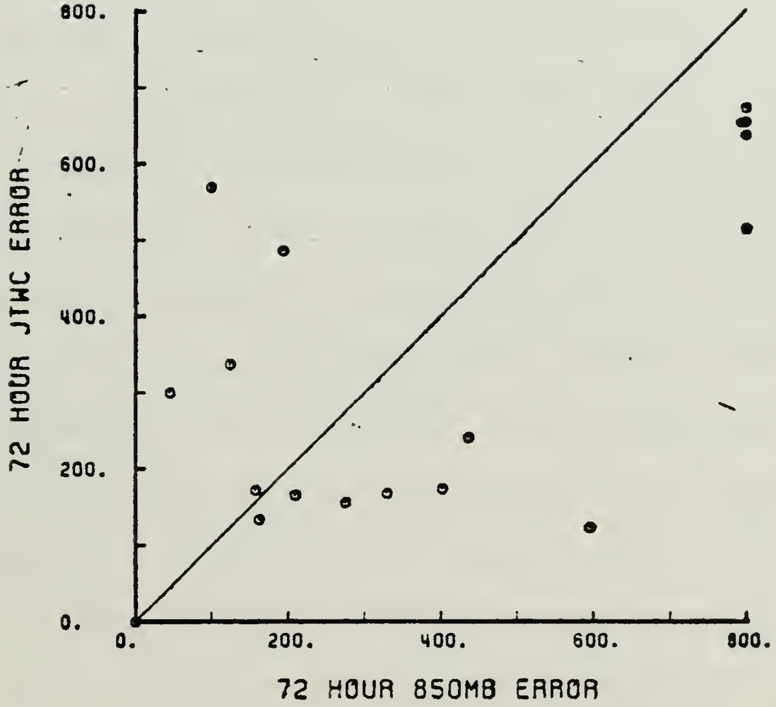


Fig. 5-9. Similar to Fig. 5-1, except the 850 mb EOF regression forecast is compared to JTWC official forecast for a 72-hour forecast.

toward large errors does not appear as dramatically in the 500mb forecasts (Figs. 5-1, 5-4 and 5-7). The superiority of the EOF forecasts to the JTWC official forecasts needs to be examined over a larger set of independent data.

One final point of interest on these figures is that both the 48-hour 850mb and 72-hour 700mb forecasts have an unusually shaped clustering of EOF regression errors at about the 150 n mi error level. No physical explanation for this clustering is known. It is very likely the event is an artifact of the data. It is, nevertheless, interesting, and worth closer examination as more data become available.

A final graphical representation of the differences in forecasting methods is shown in Figs. 5-10 through 5-12. These graphs are divided by atmospheric level, and on each are the JTWC error over the independent sample, the EOF regression forecast over the complete and homogeneous independent sample as well as the EOF forecast over the dependent sample plotted as a function of forecast time. Once again, the EOF regression scheme forecast appears superior over both the short and long term for the 500mb equations.

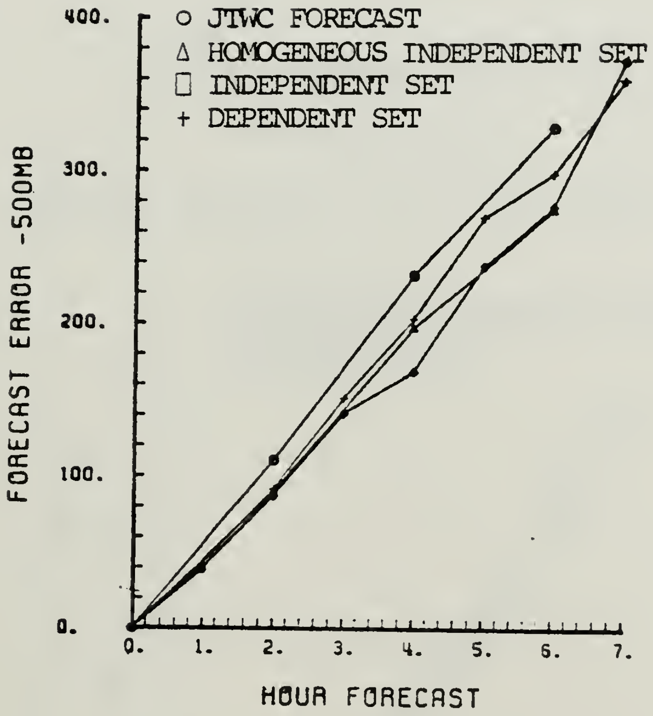


Fig. 5-10. Comparison of the JTWC official forecast over the independent data set, as well as the complete and homogeneous independent EOF regression set and the dependent set errors. All EOF results computed from 500mb equations.

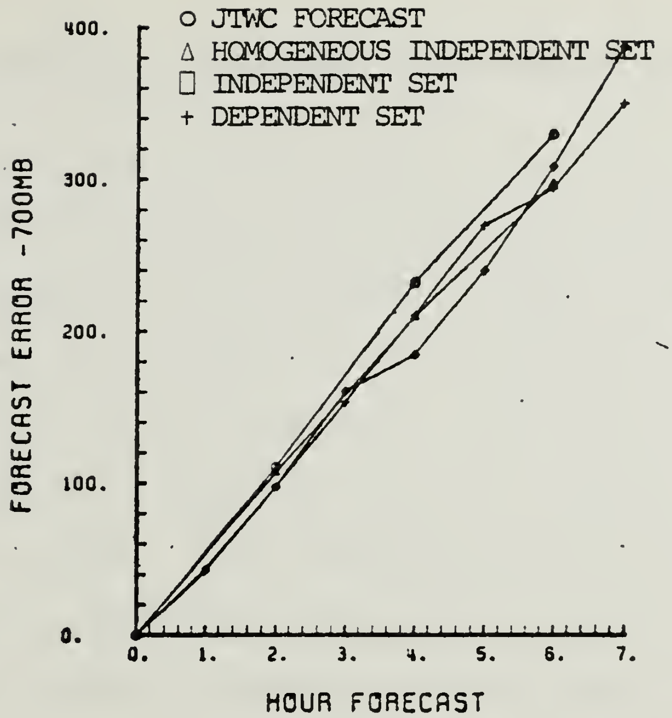


Fig. 5-11. Similar to Fig. 5-10, except EOF regression results obtained from 700mb equations.

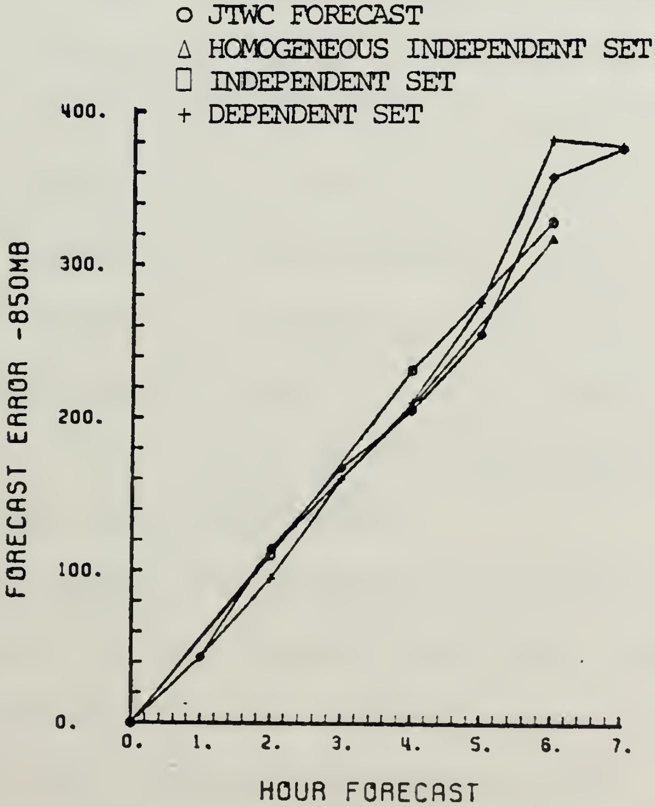


Fig. 5-12. Similar to Fig. 5-10, except EOF regression results obtained from 850mb equations.

VI. POTENTIAL FOR USE WITH INDEPENDENT DATA

Based on the results of the previous section, it appears that EOF regression forecasting has potential for improving forecasts of tropical storm movement. Using a limited independent data set, the method has been shown to be an improvement on the JTWC official forecasts. There are still unanswered questions concerning use of the model operationally on independent storms. The regression equations were derived using orthogonal coefficients derived from one set of eigenvectors. The regression equations derived are strictly valid only for tropical cyclone cases in which the coefficients are obtained from these identical vectors, so that the coefficients have a consistent meaning for each storm. If a new case is added to the dependent set, the set of vectors no longer exactly explains the maximum variation in all of the observations. Therefore, the stability of the eigenvectors and coefficients must be examined by determining whether the vectors and coefficients remain nearly the same if additional cases are added. This stability will be examined theoretically, and by a simplified experiment.

The set of dependent eigenvectors is defined as those vectors obtained from the original data set. Independent vectors are obtained from the combined set of original dependent cases plus the new independent case. If the

eigenvectors for the dependent data set are very close to the eigenvectors for the independent set, then little error will be introduced by using the dependent eigenvectors to compute the coefficients for the independent case. In this case, the independent case coefficients may be used directly in the regression equations as initially derived. If the eigenvectors are not consistent, the regression equations must be re-derived for every new forecast, including the recomputation of a new set of eigenvectors and coefficients using all data cases. Because of the large amount of computation in this case, it is highly desirable that the coefficients and vectors are consistent for independent data.

As in Chapter III, the eigenvectors are derived from solving the eigenvector equation using the known matrix $\underline{\underline{R}}$, where $\underline{\underline{R}}$ is the correlation matrix of the normalized grid points:

$$\underline{\underline{R}} = \underline{\underline{A}}\,\underline{\underline{A}}'\,N^{-1} \,. \tag{1}$$

$\underline{\underline{R}}$ is a square matrix of order equal to the number of dimensions (grid points), M. The set of eigenvectors constructed over the dependent sample should theoretically be stable if N (number of individual cases) is large. That is, addition of a single independent case should have very little effect on the shape of the observation surface in space. Inclusion of an additional data case changes $\underline{\underline{R}}$ by:

$$\underline{\underline{R}}_{NEW} = \frac{N}{N+1} \underline{\underline{R}}_{OLD} + \frac{1}{N+1} \underline{a} \; \underline{a}' , \qquad (2)$$

where $\underline{\underline{R}}_{NEW}$ is the new (independent) correlation matrix after addition of the new observation case, $\underline{\underline{R}}_{OLD}$ is the original (dependent) correlation matrix, N (N+1) the number of cases prior to (after) inclusion of the new case, and $\underline{a}$ is the (M X 1) vector of normalized D-values for the independent case. If N is initially very large, the term $\frac{1}{N+1} \underline{a} \; \underline{a}'$ in (2) is negligible compared to the first term, since the normalized observation elements are rarely greater than two or three. Therefore, to a very close approximation,

$$\underline{\underline{R}}_{NEW} \sim \underline{\underline{R}}_{OLD} , \qquad (3)$$

and the eigenvalues and vectors obtained from the dependent data should be almost identical to those obtained over all cases.

The above theory was tested with 500mb data using dependent samples of N = 50, 100, 150, 200, 300, and 400 cases with 33 independent cases. The 33 independent case orthogonal coefficients were computed in two ways:

(1) As a control, the independent case was added to the dependent sample, $\underline{\underline{R}}$ computed, and the true eigenvectors and orthogonal coefficients recalculated. Therefore, 33 separate sets of eigenvectors were computed. The eigenvectors and orthogonal coefficients are the values that minimize the deviation from the mean state for all of the data.

(2) The test method involved computing the eigenvectors only once from the dependent set (N cases). These vectors were then used to compute the orthogonal coefficients for the independent cases. If regression equations are not to be re-derived for every new operational forecast, the coefficients in the test method should be nearly identical to those from the control.

Method (2) requires considerably less computer time; however the question is whether the coefficients are sufficiently accurate. Only the first ten coefficients are examined since they represent the primary contribution to the 500mb height fields. The comparison for the first four coefficients are shown in Figs. 6-1 through 6-4. The quantity

$$Y_i = \text{ABSOLUTE VALUE } (\text{Cof}_{i_1} - \text{Cof}_{i_2}) \quad (3)$$

is summed over the 33 independent cases. Cof_{i_1} is the ith coefficient (1 to 10) computed using method (1) and Cof_{i_2} is the ith coefficient computed using method (2). The first two moments of Y_i are examined to determine the stability of the coefficients. As N increases, the standard deviations of the differences in the coefficients should become smaller.

The expected "funnel-shape" with increasing N is seen clearly in the first orthogonal coefficient (Fig. 6-1), while coefficients 2 and 3 (Figs. 6-2 and 6-3) tend to have

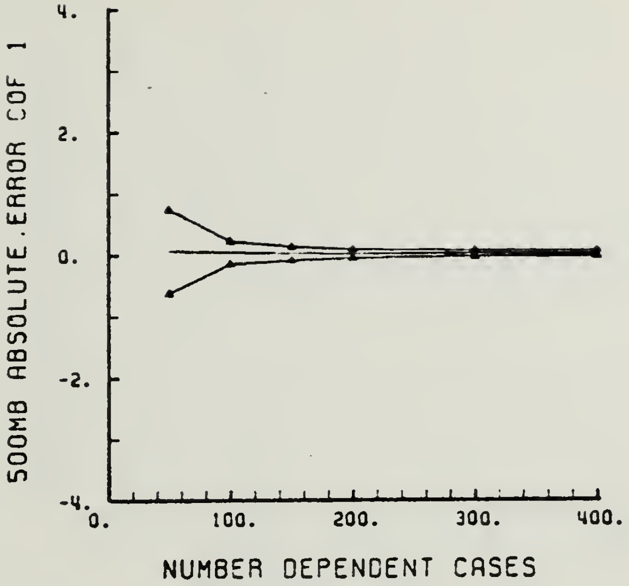


Fig. 6-1. Comparison of coefficient 1 derived over dependent and independent samples. See text for details. On the figures, the middle line is the mean and the outer two lines the 95% confidence intervals (plus/minus two standard deviations). The x-axis is the number of cases used.

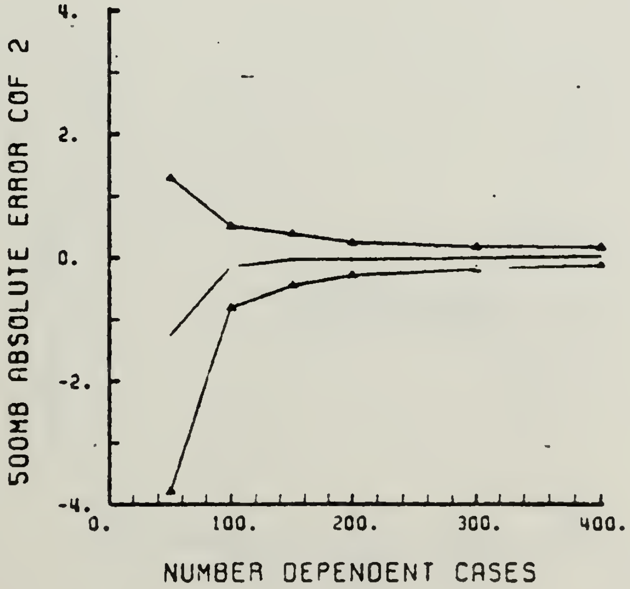


Fig. 6-2. Similar to Fig. 6-1 except for coefficient 2.

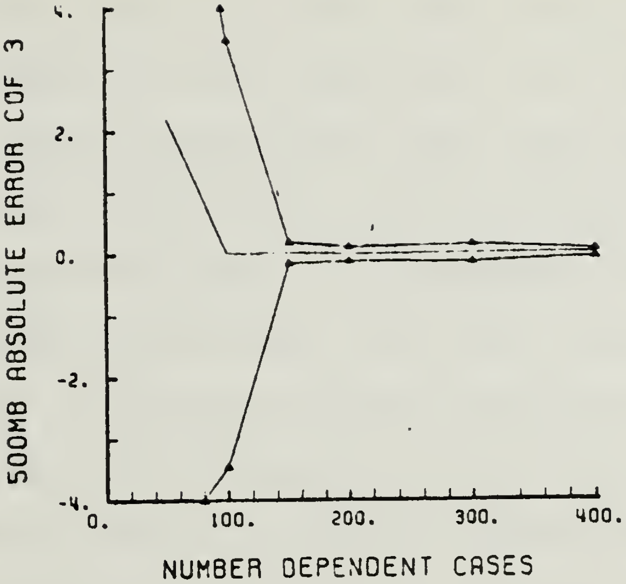


Fig. 6-3. Similar to Fig. 6-1 except for coefficient 3.

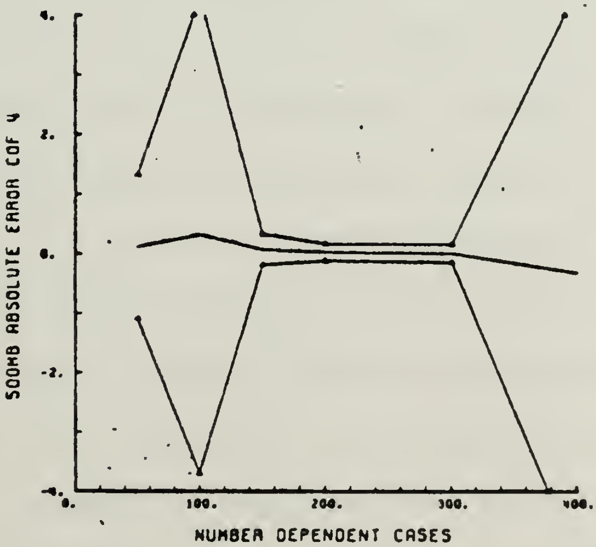


Fig. 6-4. Similar to Fig. 6-1 except for coefficient 4.

the expected shape only for N greater than 100. For the N = 50 case the mean error for both coefficients 2 and 3 is very large compared to the coefficient size (normally less than ten). This indicates the first three coefficients may be derived from the dependent set of eigenvectors determined from as few as 100 cases. An unexpected result is found with the fourth coefficient (Fig. 6-4), when N = 400 (also at N = 100). The large standard deviation indicates that at least some of the independent cases have very large error in this coefficient. A similar indication of unstable coefficients also occurs in the sixth, seventh and eighth coefficients.

The source of the error in the calculation of the coefficients was found to be due to the structure of the characteristic equation. Any single vector that is a solution eigenvector additionally represents infinite other vectors that are also solutions, and which differ only by a constant scaling factor (positive or negative). In EOF analysis, the coefficients depend upon the numerical values (and signs) of the eigenvectors. If one or two of the vectors change signs during numerical solution of the eigenvectors, then the coefficients must also reverse, which changes the EOF reconstruction. It is important to notice that the sign reversal actually occurs in deriving the new eigenvectors when the new independent case is added. In certain cases, the sign of the coefficient changes, although the magnitude

of the coefficient remains almost the same. In the cases in which some of the eigenvectors reversed signs, the error between coefficients is large. Even for these cases, the difference in the absolute values of the coefficients remains small. This is demonstrated in Fig. 6-5, in which the coefficient 4 differences are based only on the magnitude of the coefficients from the control and test methods. Large errors in the other coefficients are similarly reduced when the error differences are between absolute values of the coefficients. Once the eigenvectors and coefficients are derived from the dependent set, and the associated regression equations are generated, this set of eigenvectors must be used with any independent cases. Even though the dependent set may be quite large, the addition of a single new case will introduce the possibility of a sign change in one of the eigenvectors, and a reversal in sign of the coefficients. This would invalidate the original regression equation set, and require a re-derivation of both the eigenvectors and the regression equations with each new entry into the sample.

The reversal in sign of the coefficients and vectors is probably due to computer round-off error. Solution of a 120 dimension eigenvalue problem requires simultaneous solution of 120 homogeneous equations--which is an extremely ill-conditioned problem (Gerald, 1977). The probability of catastrophic round-off error increases dramatically as the

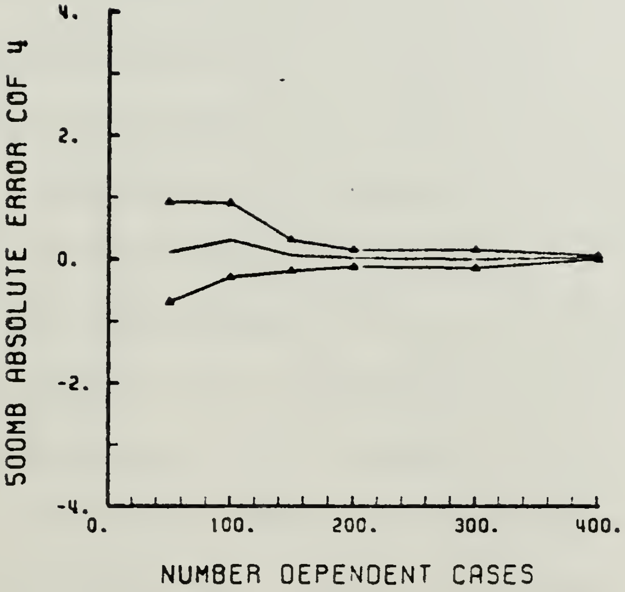


Fig. 6-5. Similar to Fig. 6-1 except for coefficient 4 using absolute differences of the derived coefficients only.

number of dimensions increase. However, this reversal problem is not significant in the study, as long as the coefficients for independent cases are calculated from dependent eigenvectors.

Further attempts to isolate the conditions under which this reversal occurs were without success. Random tests were conducted in 3, 5, 9 and 20 dimensions. Not until dimension size reached 20 were the first reversals noticed. The fact that the reversal does not occur until higher dimension systems are used is consistent with the argument above, because the greater the number of dimensions, the greater the probability for catastrophic round-off error.

Because the coefficients calculated by the two methods have consistent magnitudes, it may be concluded that the coefficients computed for independent cases using the same dependent eigenvectors will introduce very little error to the movement forecast. Thus, implementation of these EOF regression forecasts with independent cases becomes straightforward. Only two major operations are required. First, the EOF orthogonal coefficients from the dependent set of eigenvectors are stored. This involves multiplication of a (10 X 120) transpose matrix of truncated eigenvectors and the (120 X 1) normalized observation vector, which gives the ten coefficients. The second step involves simple substitution of the independent coefficients into the regression equations. The same eigenvectors and eigenvalues

may be used indefinitely on independent storms, although it is recommended the regression equations be updated at the conclusion of each typhoon season.

VII. CONCLUSIONS AND FUTURE APPLICATIONS

It has been shown that EOF coefficients correlate strongly with the observed motion. Therefore, use of EOF coefficients to represent the geopotential patterns in the environment of a tropical cyclone appears to be a valid approach for incorporation of synoptic information into a statistically based forecast. Incorporation of synoptic forcing by using EOF coefficients appears to have potential in forecasting tropical storm motion. Using an independent sample, an average of 17% improvement relative to JTWC official motion forecasts was obtained using the 500mb EOF regression equations. The use of 500mb equations gave better forecasts than either the 700mb or 850mb equations. In contrast, Brown (1981) found no significant difference in forecast ability in a map-typing forecast technique using the same three atmospheric levels. Since this is only a pilot study, the good results shown here need to be tested further with new data cases. Several conclusions and future applications are drawn from this study.

(1) The regression equations were developed with a fairly small dependent data sample, and yet gave good results when tested with an independent sample. As the number of useable storm cases for the dependent sample increases, the regression equations should become progressively more refined. As

the dependent data size increases, in any regression scheme, more extreme cases are typically forecast better. Large forecast errors should occur less frequently with a larger data sample.

(2) This method of incorporating synoptic fields into the regression equations is not limited to observed fields. It is likely that coefficients derived from a 24-hour forecast field (from dynamic numerical weather prediction models) would improve the long range forecast. As seen in the study, the accuracy of the regression equations decreased sharply in time. This study used only the current observed field. After 24 to 36 hours, it is expected that the forcing from the mid-latitudes would be significantly different. Use of a 24 hour prognosis field might give a better representation of the forcing in the long-range forecast.

(3) The model is extremely simple. Using only values representing the synoptic forcing in a limited grid region about the storm, past storm movement and an intensity measure (which proved to be of little value), the forecasts appear to be very good. If variables representing other physical features thought to impact storm movement are incorporated into the regression equations, even better forecasts should be possible. It is possible that the phase of equatorial planetary waves near the storm, and other large scale circulation features may play a role in tropical storm movement. These waves are not easily detected.

Holton (1972) notes that these waves are usually only identifiable in the stratosphere, although they extend throughout the troposphere and stratosphere. It is possible that these waves could be identified using an EOF analysis of the global band in the tropics at a mid-tropospheric level. For instance, a global tropical grid, with coverage to about 30°N and 30°S may be adequate to identify these waves (which would probably be seen in the first 5 to 10 eigenvectors). These EOF coefficients could then be incorporated into the regression equation. A global grid could also possibly detect features such as the Walker circulation, and these features could be incorporated into the regression forecast. A better storm intensity than the maximum wind used in this study needs to be found. Variables such as the radius of maximum winds should be tested as the data become available. The potential predictors that could be included are certainly not limited to those mentioned above.

(4) The model was developed for use in the western North Pacific Ocean genesis basin, although the method could be developed for other genesis regions. The only difference in the different regions would be in the values of the regression coefficients.

(5) Rotation of eigenvectors could also be tried to improve the model. If this were to be done, the number of

retained vectors would have to be larger, to prevent against underfactoring.

(5) Application of the EOF scheme in its present form would be a simple matter. In fact, if the regression equations were updated only once a year, the entire forecast could conceivably be obtained on a hand-held programmable calculator with sufficient memory to store the mean and standard deviation of the grid points and all eigenvectors. Entry of the data at the 120 grid points is all that would be required to generate the movement forecast. The grid point data might be obtained using a Bessel linear interpolation from the 63 X 63 FNOC analysis. Therefore, the scheme could be implemented for operational use with a minimum effort.

In conclusion, the EOF regression scheme shows great promise for improvement of operational forecasts of tropical storm movement. In this pilot study, using a very simple model, the scheme performed very well. Potential improvement is possible through addition of more sophisticated physical forcing parameters and forecast dynamic fields that may affect storm movement. Further research in this area is definitely warranted.

APPENDIX A

700 AND 850MB EIGENVECTORS

The first 10 eigenvectors for the 700 and 850mb level follow. These are the vectors used in deriving the coefficients used in the regression equations.

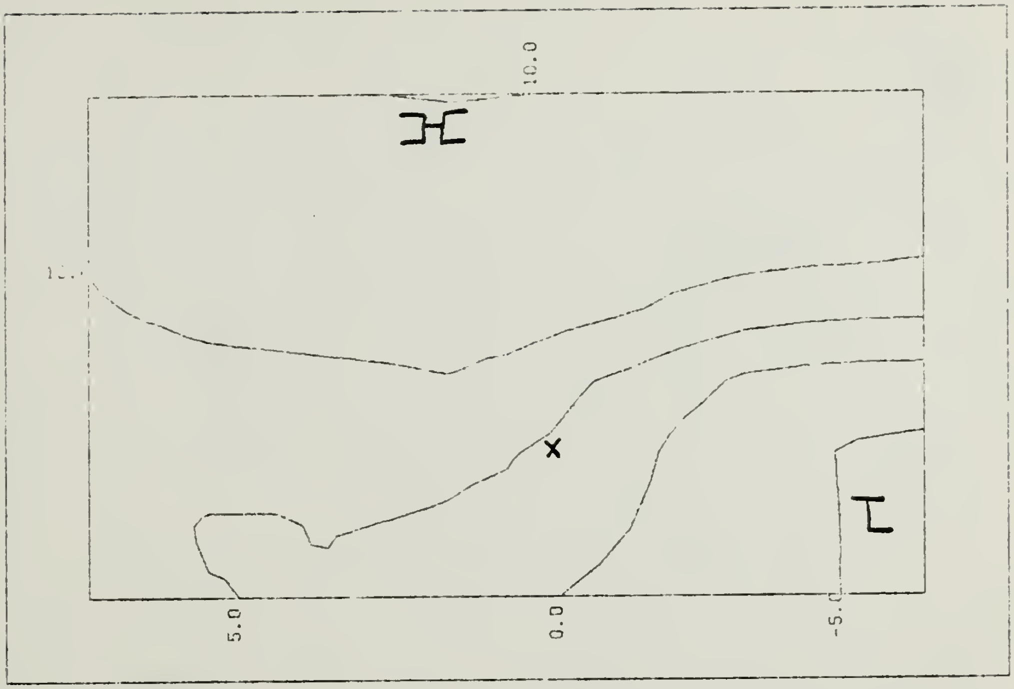


Fig. A1-1. Eigenvector 1 elements (multiplied by 100) at 700mb with the tropical cyclone located at the x-position.

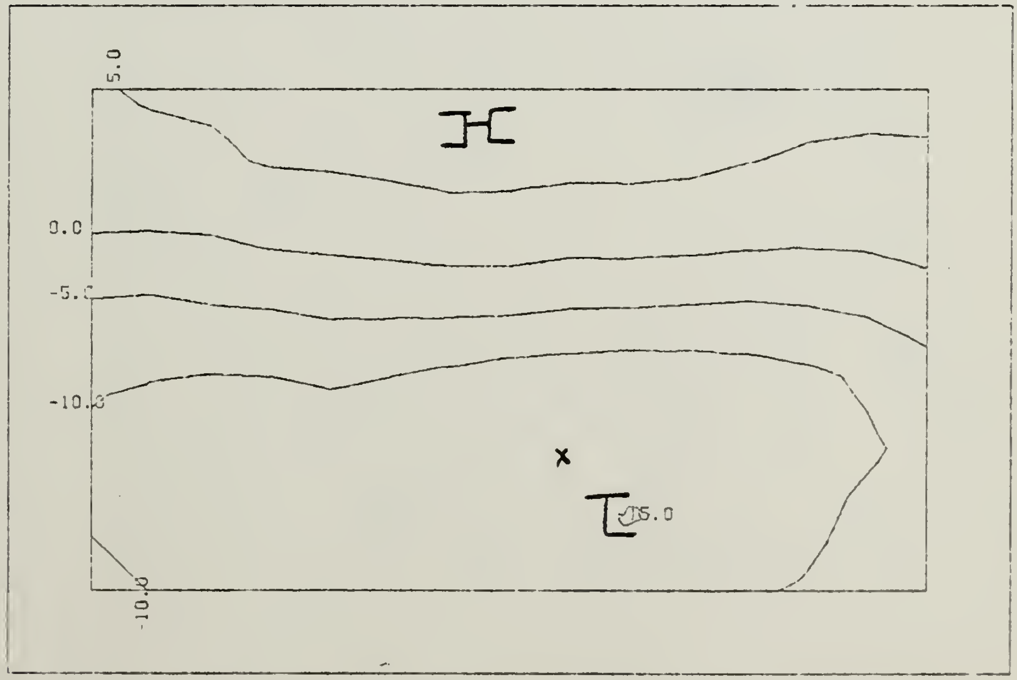


Fig. A1-2. Similar to Fig. A1-1 except for eigenvector 2.

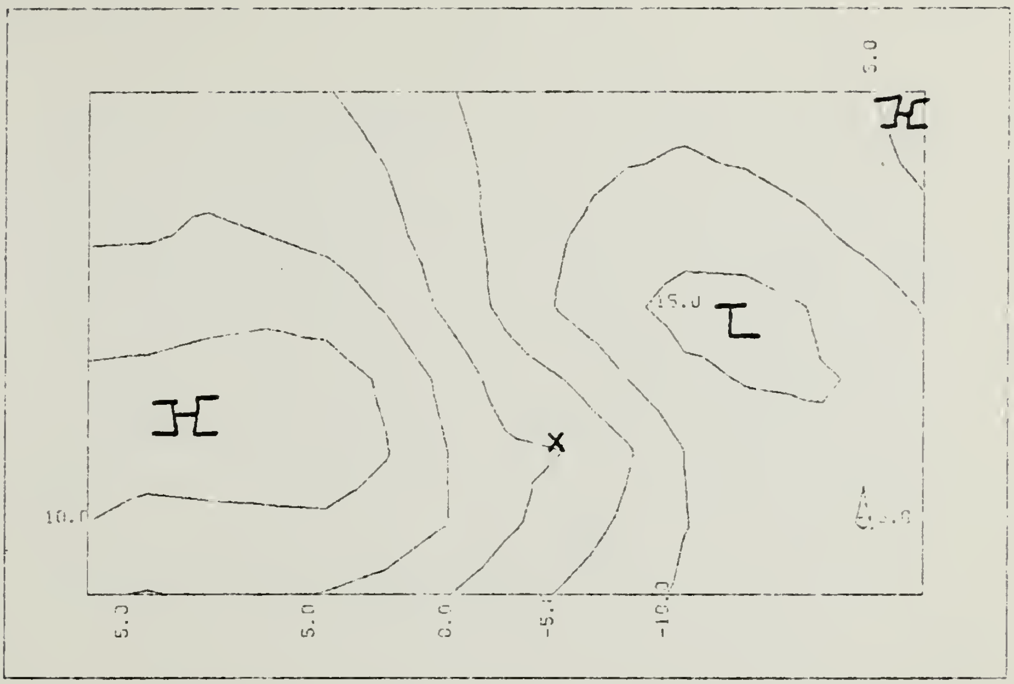


Fig. Al-3. Similar to Fig. Al-1 except for eigenvector 3.

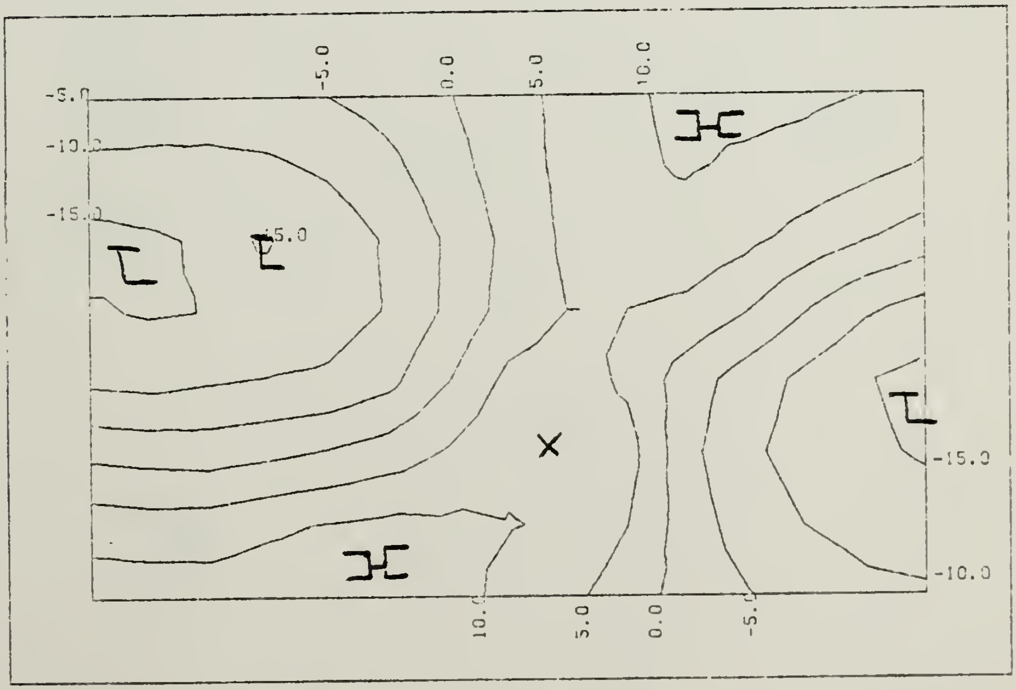


Fig. Al-4. Similar to Fig. Al-1 except for eigenvector 4.

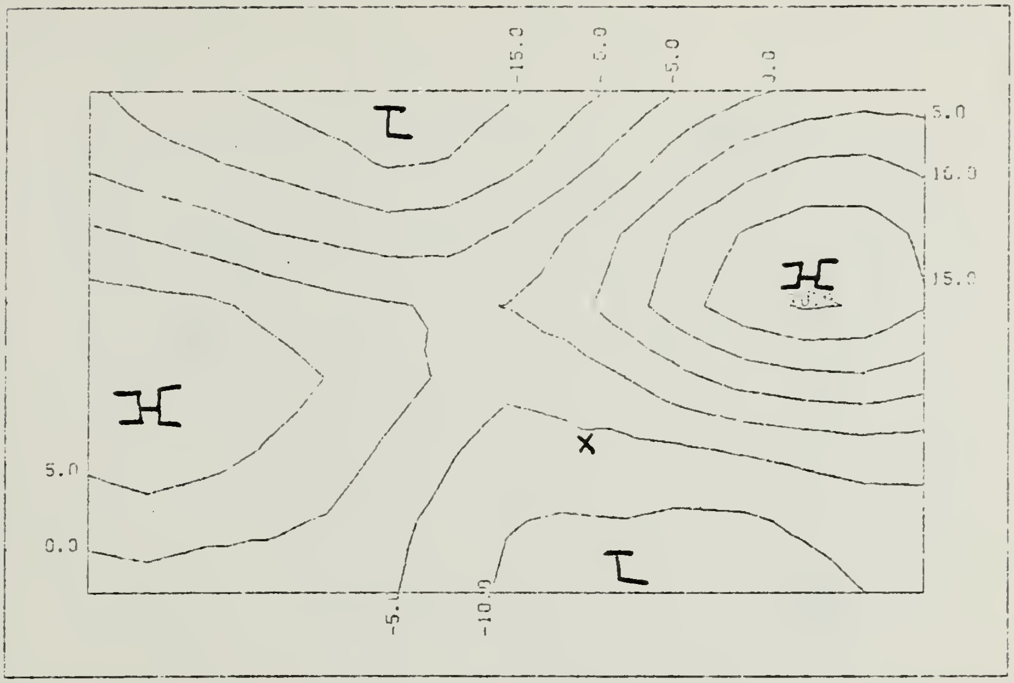


Fig. A1-5. Similar to Fig. A1-1 except for eigenvector 5.

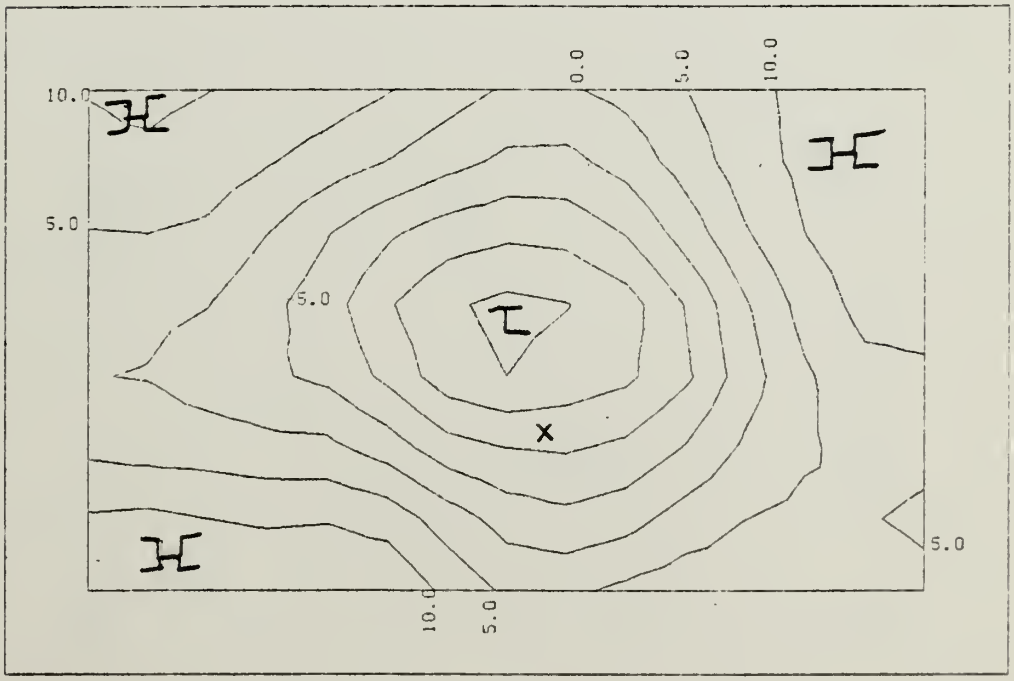


Fig. A1-6. Similar to Fig. A1-1 except for eigenvector 6.

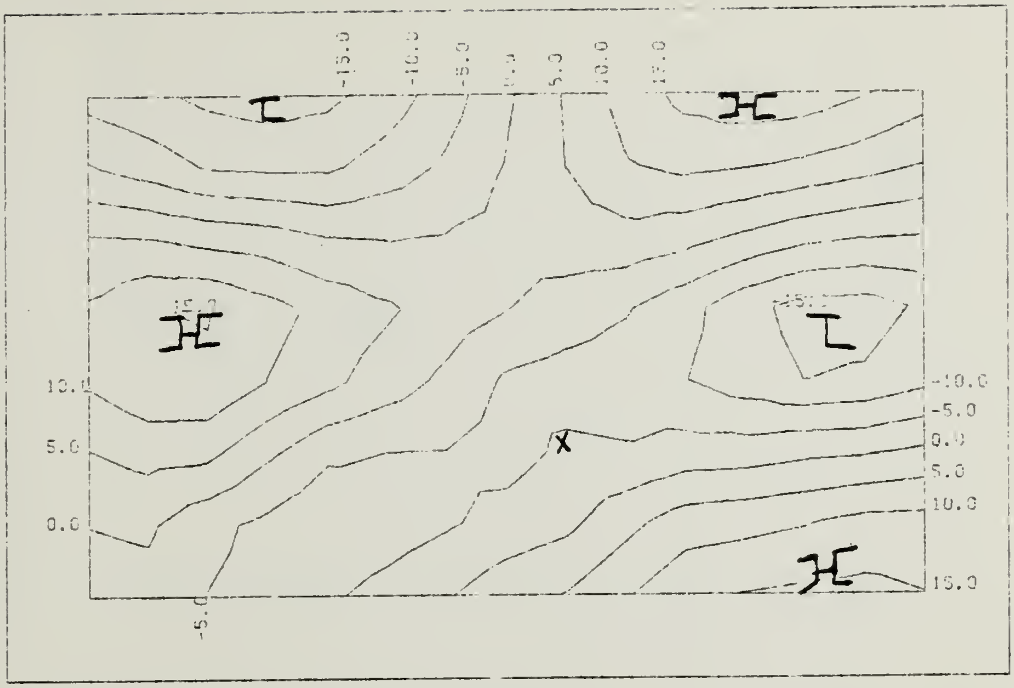


Fig. A1-7. Similar to Fig. A1-1 except for eigenvector 7.

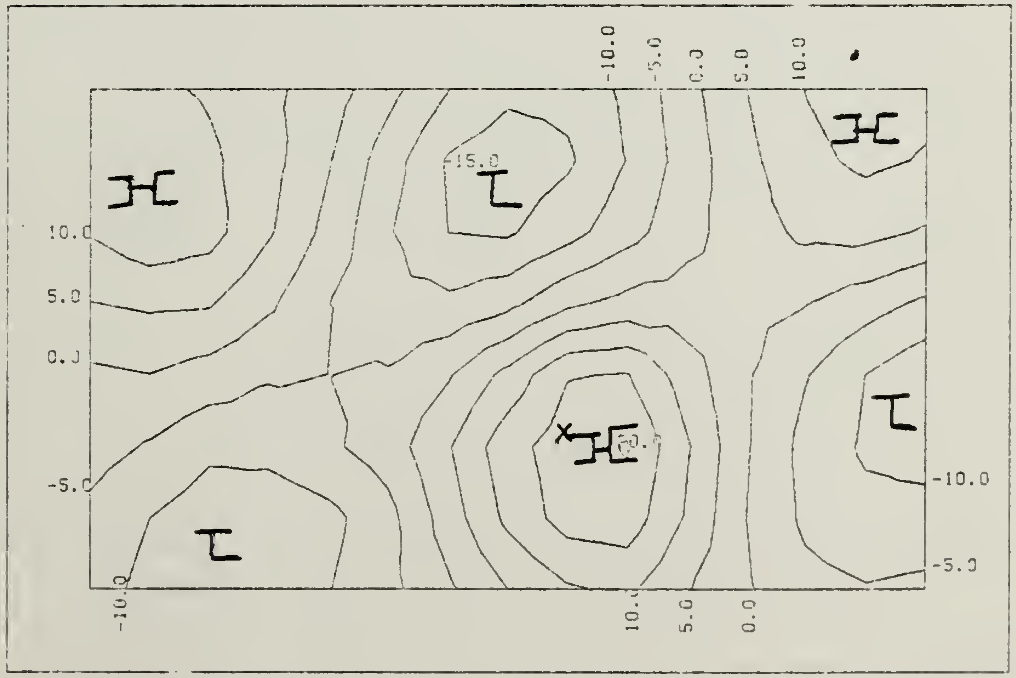


Fig. A1-8. Similar to Fig. A1-1 except for eigenvector 8.

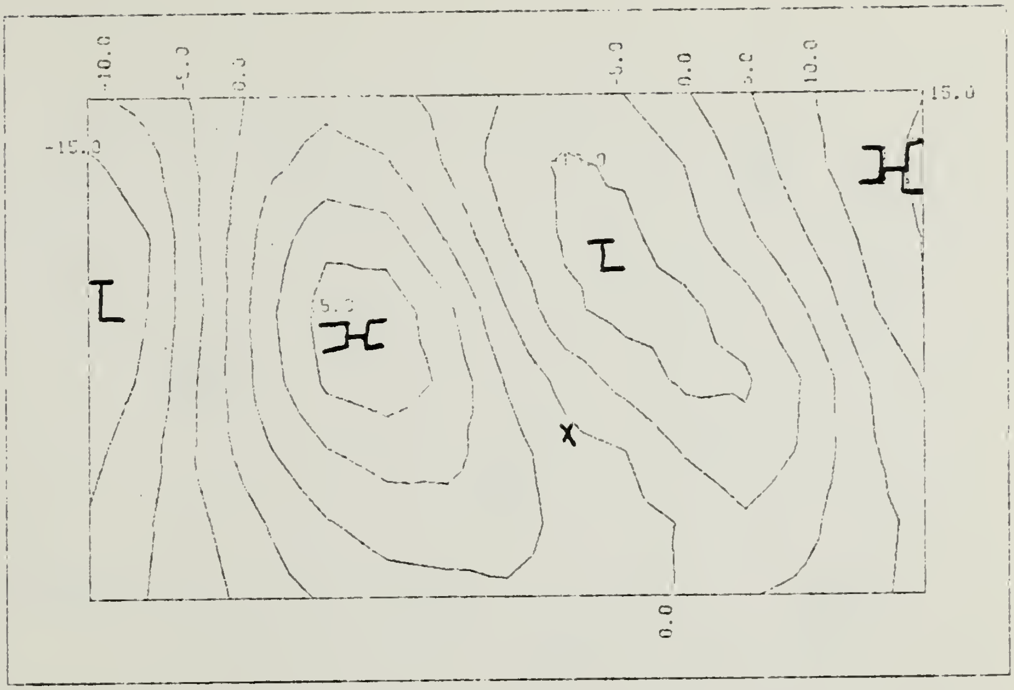


Fig. A1-9. Similar to Fig. A1-1 except for eigenvector 9.

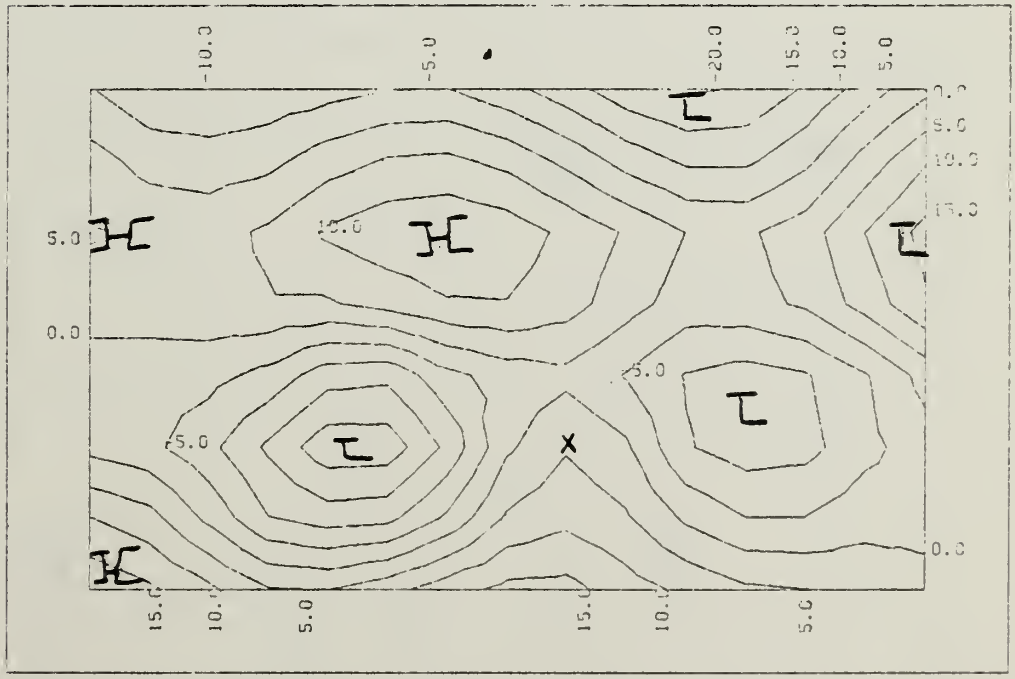


Fig. A1-10. Similar to Fig. A1-1 except for eigenvector 10.

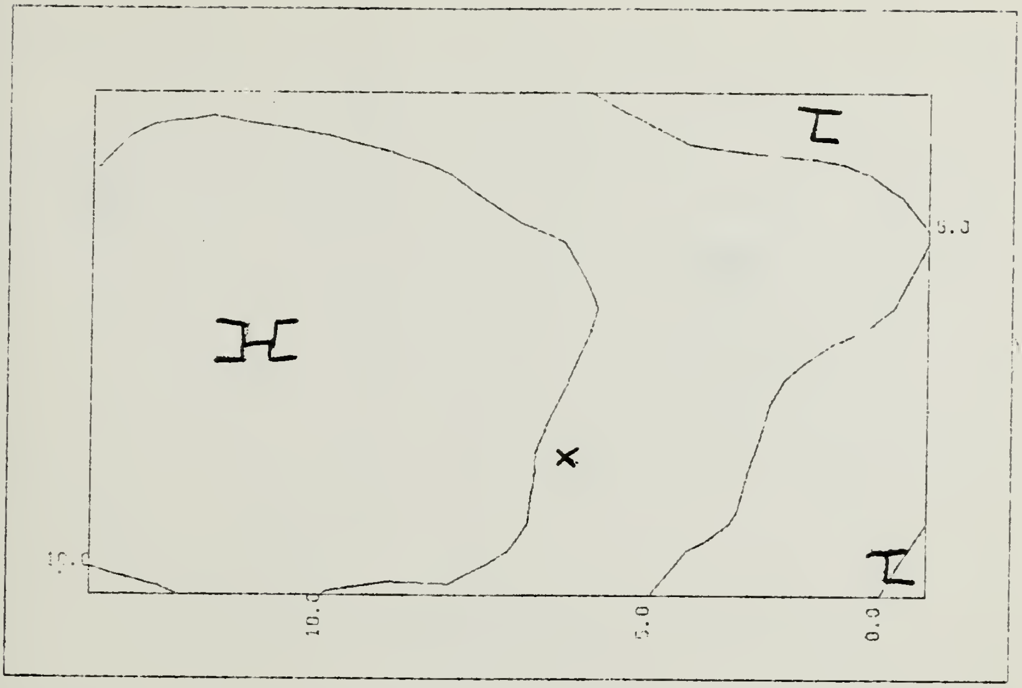


Fig. A1-11. Similar to Fig. A1-1 except for 850mb level.

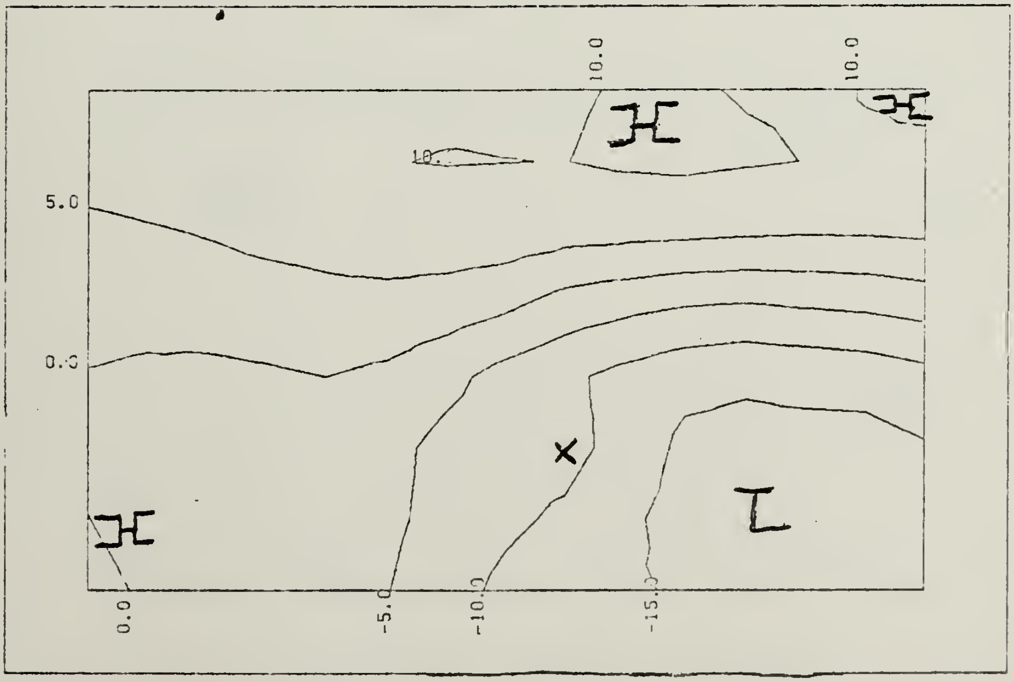


Fig. A1-12. Similar to Fig. A1-11 except for eigenvector 2.

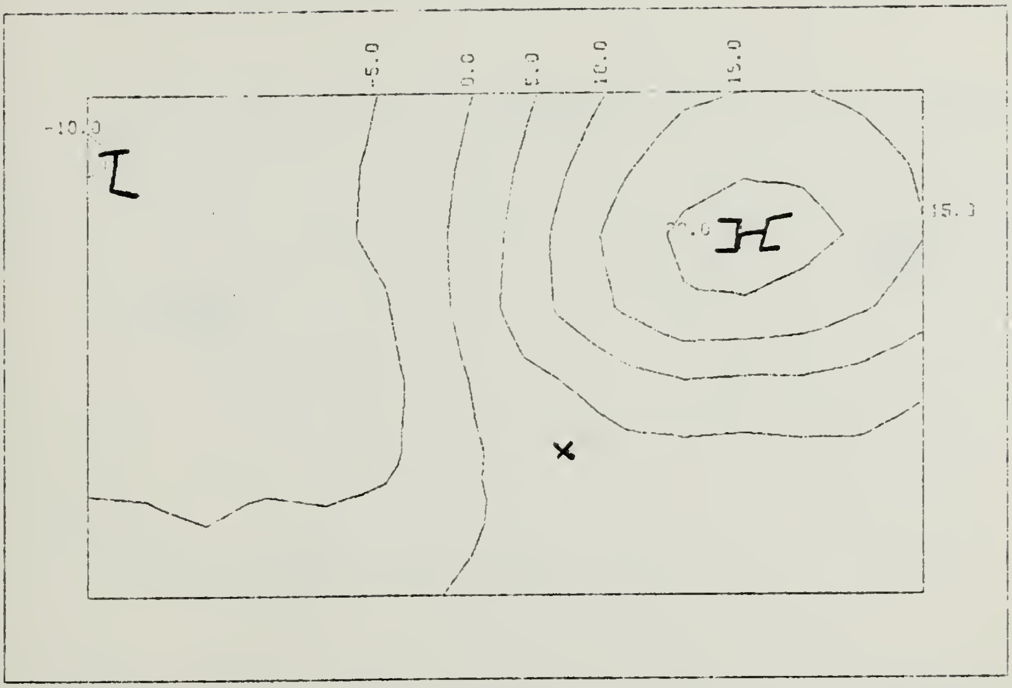


Fig. A1-13. Similar to Fig. A1-11 except for eigenvector 3.

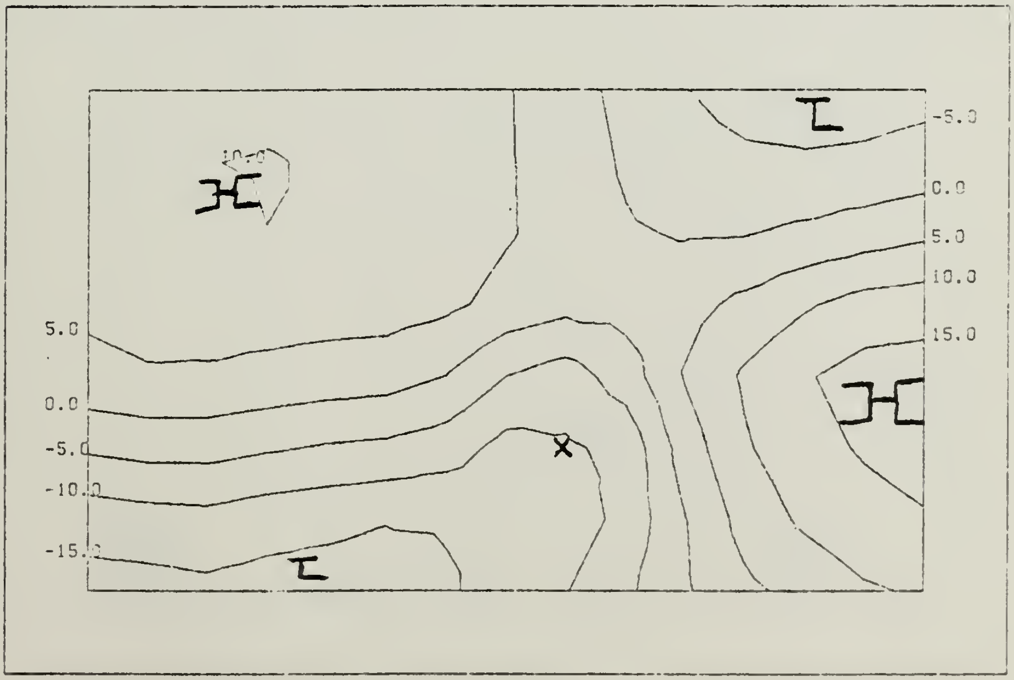


Fig. A1-14. Similar to Fig. A1-11 except for eigenvector 4.

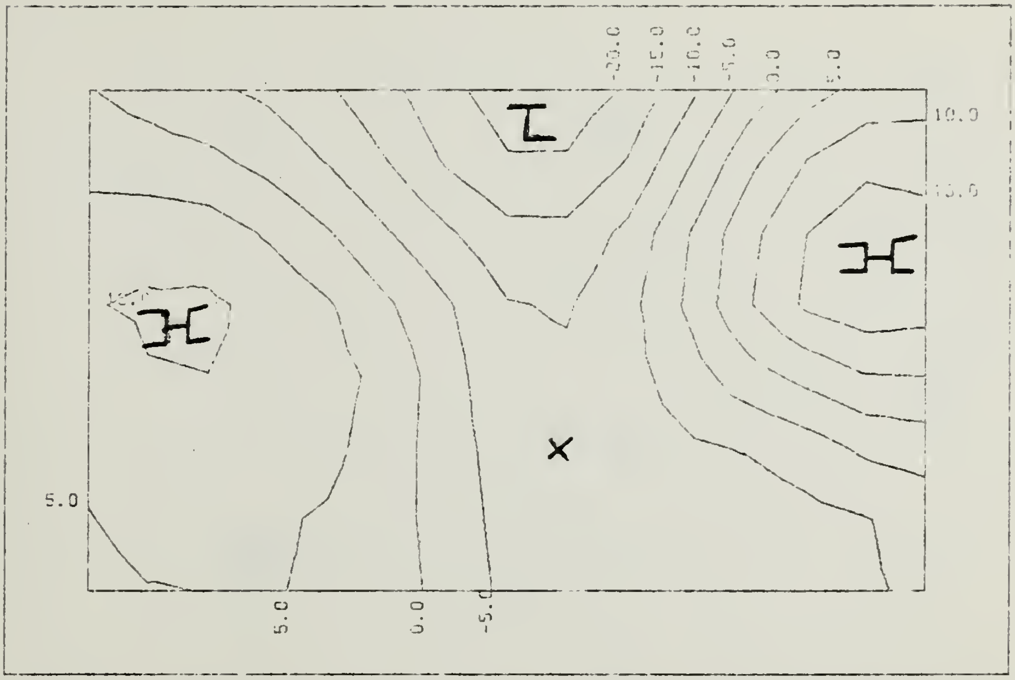


Fig. A1-15. Similar to Fig. A1-11 except for eigenvector 5.

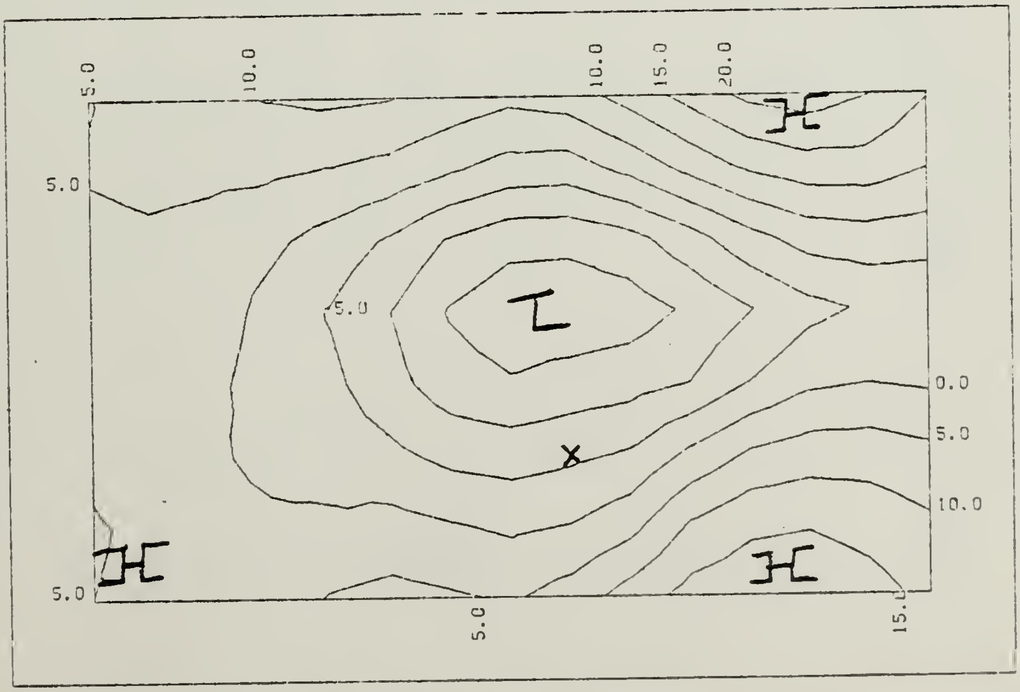


Fig. A1-16. Similar to Fig. A1-11 except for eigenvector 6.

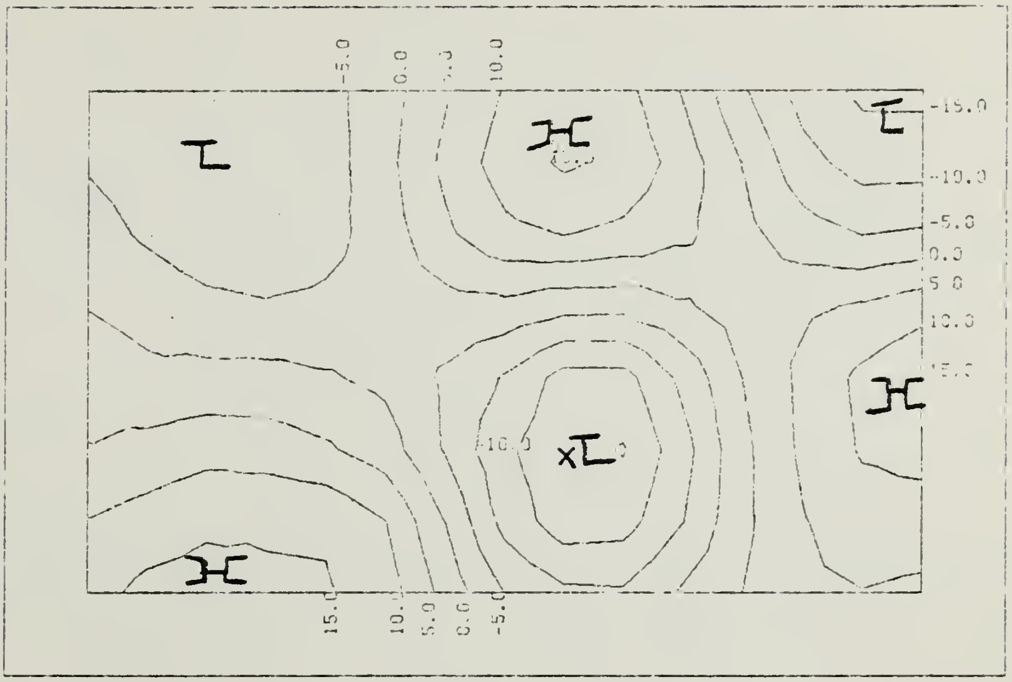


Fig. A1-17. Similar to Fig. A1-11 except for eigenvector 7.

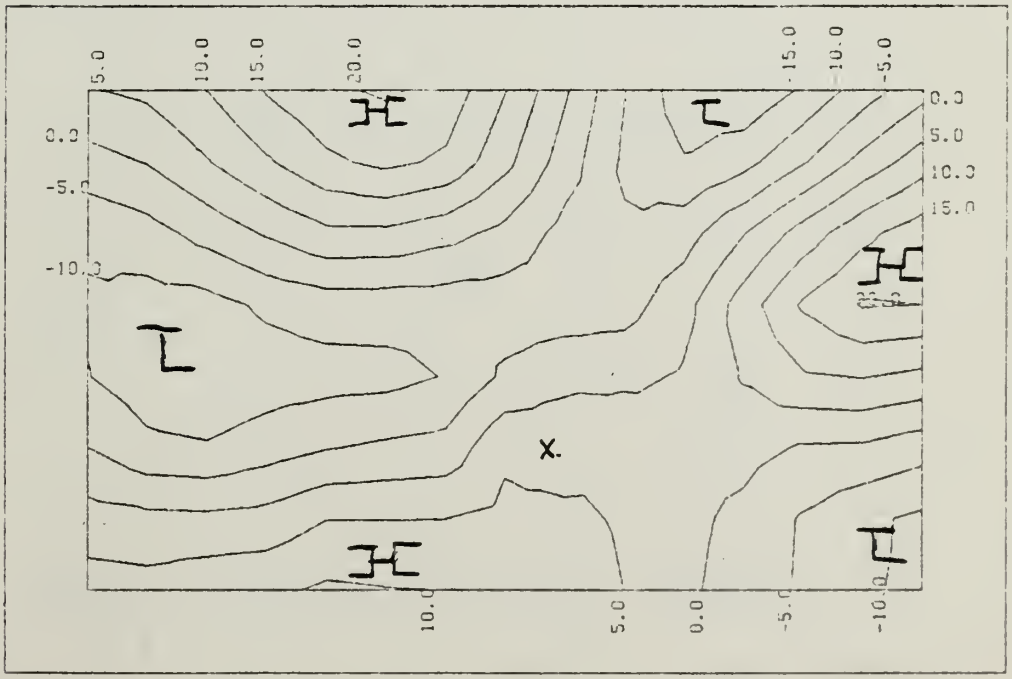


Fig. A1-18. Similar to Fig. A1-11 except for eigenvector 8.

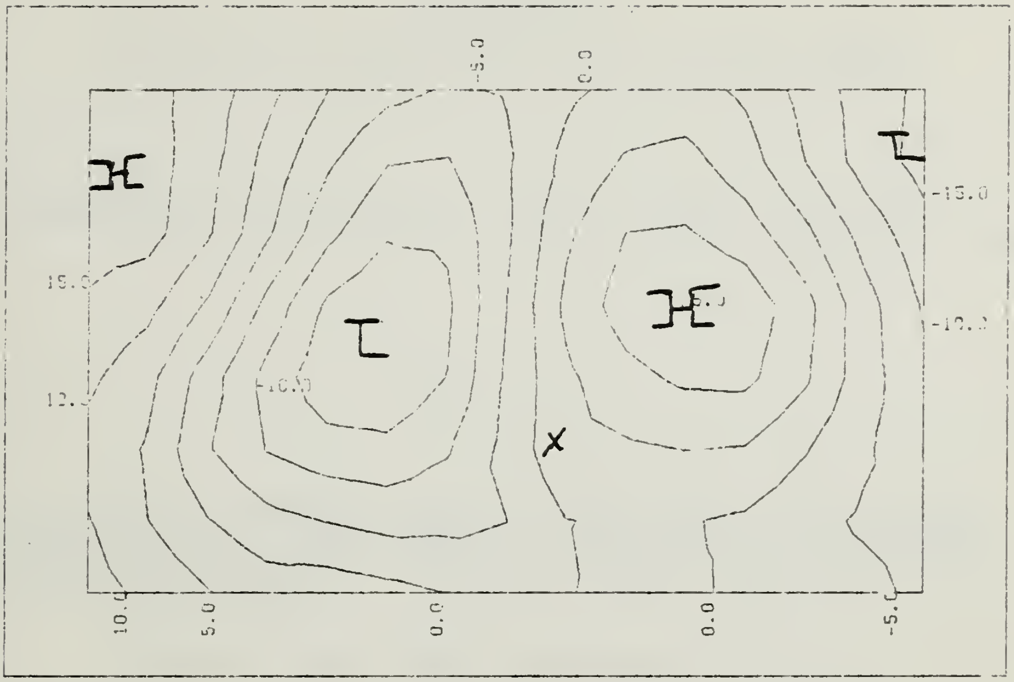
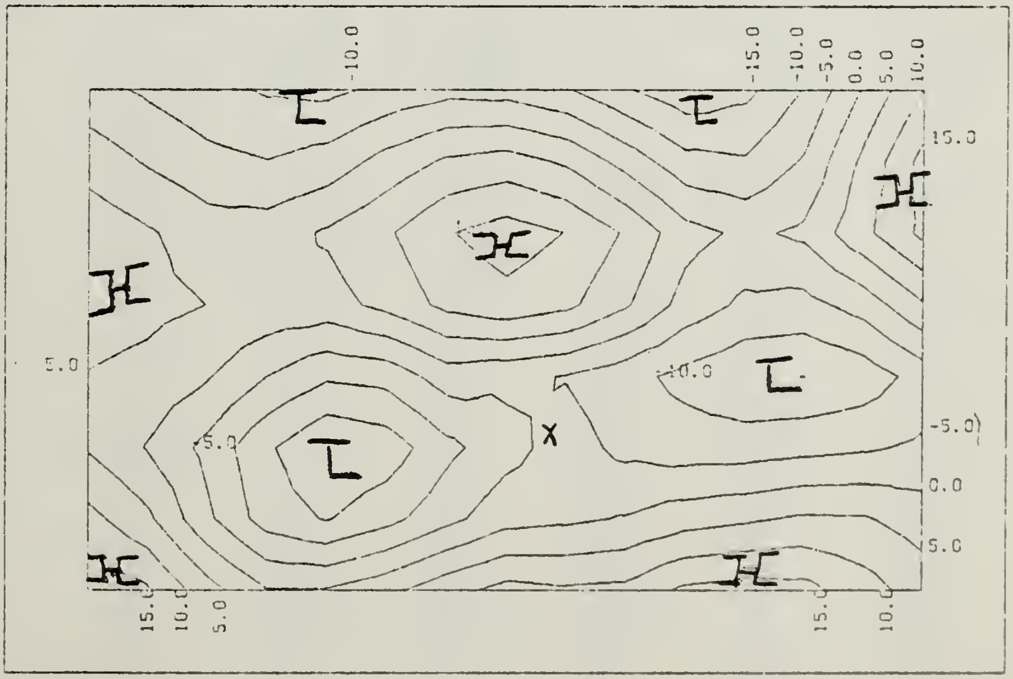


Fig. A1-19. Similar to Fig. A1-11 except for eigenvector 9.



. Fig. A1-20. Similar to Fig. A1-11 except for eigenvector 10.

APPENDIX B

REGRESSION COEFFICIENTS FOR 700 AND 850MB

The regression coefficients for the 700 and 850mb equations follow.

TABLE B - 1

Regression coefficients for the seven meridional equations using 700mb EOF's. A value of .0 indicates the predictor was not selected in the stepwise selection procedure.

FORECAST VALID FOR BASE TIME PLUS HOURS

	12	24	36	48	60	72	84
Intercept	33.485	55.890	101.502	113.129	226.841	257.266	340.398
Cof1	.0	-1.784	-3.988	.0	.0	.0	19.739
Cof2	-1.630	-3.775	-6.681	-3.812	-7.215	-7.605	.0
Cof3	-4.007	-7.917	-11.190	-14.186	-17.901	-22.005	.0
Cof4	.0	.0	.0	.0	.0	.0	.0
Cof5	.0	.0	.0	.0	.0	.0	.0
Cof6	4.205	10.826	16.118	9.909	13.959	.0	.0
Cof7	.0	4.238	7.284	17.866	24.736	32.436	.0
Cof8	.0	.0	.0	.0	.0	.0	.0
Cof9	-5.307	-10.641	-15.707	.0	.0	.0	.0
Cof10	.0	-5.012	-9.934	.0	.0	.0	.0
Plat1	.0	0.713	0.824	1.279	1.731	1.463	.0
Plat2	.389	.0	.0	.0	.0	.0	0.569
Plat3	.0	.0	.0	.0	.0	.0	.0
Plat4	-0.306	.0	.0	.0	.0	.0	.0
Plat5	.0	.0	.0	.0	.0	.0	.0
Plat6	.0	.0	.0	.0	-0.595	.0	.0
Plon1	0.222	.0	.0	.0	.0	.0	.0
Plon2	-0.098	.0	.0	.0	.0	.0	.0
Plon3	.0	.0	.0	.0	.0	.0	.0
Plon4	.0	.0	.0	.0	.0	.0	.0
Plon5	.0	.0	.0	.0	.0	.0	.0
Plon6	.0	.0	.0	.0	.0	.0	.917
Amw0	.0	.0	.0	.0	.0	.0	.0
Amw1	.0	0.317	0.493	0.130	.0	.0	.0
Amw2	.0	.0	.0	.0	.0	.0	.0
Amw3	.0	.0	.0	.0	.0	.0	.0

TABLE B - 2

Regression coefficients for the seven zonal equations using 700mb EOF's. A value of .0 indicates the predictor was not selected in the stepwise selection procedure.

FORECAST VALID FOR BASE TIME PLUS HOURS

	12	24	36	48	60	72	84
Intercept	28.246	52.181	88.091	83.395	206.607	333.156	382.634
Cof1	1.759	4.857	11.116	16.472	19.922	23.270	24.606
Cof2	.0	2.463	4.558	7.466	12.687	.0	.0
Cof3	-2.618	-6.010	-10.575	-10.074	-16.974	-18.865	-29.982
Cof4	2.270	4.774	7.753	.0	.0	.0	.0
Cof5	3.066	5.875	11.750	15.589	26.300	45.822	51.012
Cof6	-2.451	.0	-8.754	.0	.0	-24.452	-28.619
Cof7	.0	.0	.0	.0	.0	.0	.0
Cof8	-3.301	-5.380	-10.965	.0	-21.056	-31.349	-50.614
Cof9	-3.268	.0	.0	.0	.0	.0	.0
Cof10	.0	.0	.0	.0	.0	.0	.0
Plat1	.0	.0	.0	-0.797	-1.000	.0	.0
Plat2	.0	.0	.0	.0	.0	-0.787	.0
Plat3	.0	.0	-0.159	.0	.0	.0	.0
Plat4	.0	.0	.0	.0	.0	.0	-1.390
Plat5	.0	.0	.0	.0	.0	.0	.0
Plat6	-0.124	-0.261	.0	.0	.0	.0	.0
Plon1	-0.578	-1.486	-1.853	-2.164	-2.325	-2.308	-2.051
Plon2	-0.083	.0	.0	.0	.0	.0	.0
Plon3	.0	.0	.0	.0	.0	.0	.0
Plon4	.0	.0	.0	.0	.0	.0	.0
Plon5	.0	.0	.0	.0	.0	.0	.0
Plon6	.0	.0	.0	.0	.0	.0	.0
Amw0	.0	.0	.0	.0	.0	.0	.0
Amw1	.0	.0	.0	.0	.0	.0	.0
Amw2	.0	.0	.0	.0	.0	-2.259	-2.206
Amw3	-0.190	-0.400	-0.611	.0	-1.550	.0	.0

TABLE B - 3

Regression coefficients for the seven meridional equations using 850mb EOF's. A value of .0 indicates the predictor was not selected in the stepwise selection procedure.

FORECAST VALID FOR BASE TIME PLUS HOURS

	12	24	36	48	60	72	84
Intercept	26.555	55.682	77.256	121.233	211.106	324.600	207.533
Cof1	.0	.0	.0	.0	.0	.0	.0
Cof2	1.154	3.641	7.988	11.981	16.514	11.960	.0
Cof3	3.865	9.081	17.534	19.859	31.471	13.913	38.864
Cof4	.0	.0	.0	.0	.0	33.760	.0
Cof5	.0	.0	.0	.0	.0	.0	.0
Cof6	-3.117	.0	.0	.0	-24.926	22.221	.0
Cof7	2.812	.0	.0	.0	.0	-45.311	41.016
Cof8	.0	.0	.0	.0	.0	26.327	.0
Cof9	3.894	9.170	.0	.0	.0	.0	.0
Cof10	.0	.0	.0	.0	.0	.0	.0
Plon1	.0	.0	-0.404	-0.723	.0	.0	.0
Plon2	.0	.0	.0	.0	.0	.0	.0
Plon3	.0	.0	.0	.0	-0.358	-0.764	.0
Plon4	.0	.0	.0	.0	.0	.0	-1.470
Plon5	.0	.0	.0	.0	.0	.0	.0
Plon6	-0.114	-0.331	.0	.0	.0	.0	.0
Plat1	-0.593	-1.542	-2.147	-2.477	-2.457	.0	-2.992
Plat2	-0.089	.0	.0	.0	.0	-2.409	.0
Plat3	.0	.0	.0	.0	.0	.0	.0
Plat4	.0	.0	.0	.0	.0	.0	.0
Plat5	.0	.0	.0	.0	.0	.0	.0
Plat6	.0	.0	.0	.0	.0	.0	.0
Amw0	.0	.0	.0	.0	.0	.0	.0
Amw1	.0	.0	.0	.0	.0	.0	.0
Amw2	.0	.0	.0	.0	.0	.0	.0
Amw3	-0.208	-0.481	-0.794	-0.856	-1.697	-2.407	.0

TABLE B - 4

Regression coefficients for the seven zonal equations using 850mb EOF's. A value of .0 indicates the predictor was not selected in the stepwise selection procedure.

	FORECAST VALID FOR BASE TIME PLUS HOURS						
	12	24	36	48	60	72	84
Intercept	29.935	72.723	92.290	158.753	210.892	190.344	309.226
Cof1	.0	.0	.0	.0	.0	.0	.0
Cof2	-2.286	-5.569	-9.653	-7.213	-9.184	-15.096	.0
Cof3	2.383	4.675	.0	11.292	15.976	14.822	13.543
Cof4	.0	.0	.0	.0	.0	.0	.0
Cof5	1.886	4.859	.0	.0	.0	.0	.0
Cof6	4.692	11.561	18.413	17.100	23.592	31.160	24.811
Cof7	.0	5.729	9.021	9.773	.0	.0	.0
Cof8	.0	.0	.0	.0	.0	.0	.0
Cof9	4.569	7.327	9.740	.0	.0	.0	.0
Cof10	.0	.0	.0	.0	.0	.0	.0
Plon1	.0	0.807	1.011	1.374	1.558	1.280	1.026
Plon2	0.393	.0	.0	.0	.0	.0	.0
Plon3	.0	.0	.0	.0	.0	.0	.0
Plon4	-0.272	.0	.0	.0	.0	.0	.0
Plon5	.0	.0	.0	.0	.0	.0	.0
Plon6	.0	.0	.0	.0	.0	.0	.0
Plat1	.0	.0	.0	.0	.0	.0	.0
Plat2	.0	0.192	.0	.0	.0	.0	.0
Plat3	.0	.0	.0	.0	.0	.0	.0
Plat4	.0	-0.415	.0	.0	.0	.0	.0
Plat5	.0	.0	.0	.0	.0	.0	.0
Plat6	.0	.0	.0	.0	.0	.0	.0
Amw0	.0	.0	.0	.0	.0	.0	.0
Amw1	.0	.0	0.486	.0	.0	1.134	.0
Amw2	.0	.0	.0	.0	.0	.0	.0
Amw3	.0	.0	.0	.0	.0	.0	.0

APPENDIX C
MODIFIED REGRESSION EQUATION RESULTS

The enclosed table gives the R^2 statistic, and the sample size for each atmospheric level, for the modified regression equations. These equations were derived using only 13 potential predictors, the 10 coeffiecients, Plat1, Plon1 and Amw1. The values may be compared with Table 5-3 using the entire set of 26 predictors.

TABLE C - 1

Sample size and R^2 statistic for each zonal and meridional modified regression equation by forecast time and atmospheric level.

	FORECAST INTERVAL (HR)						
	12	24	36	48	60	72	84
NUMBER OF DEPENDENT DATA CASES	409	409	387	307	281	203	184
ZONAL EQUATIONS							
500mb	.777	.714	.672	.594	.549	.519	.457
700mb	.758	.695	.649	.574	.544	.541	.470
850mb	.738	.676	.614	.536	.497	.503	.456
MERIDIONAL EQUATIONS							
500mb	.483	.441	.395	.325	.229	.252	.169
700mb	.465	.435	.378	.315	.228	.202	.145
850mb	.431	.396	.337	.285	.225	.219	.111

LIST OF REFERENCES

Anderson, T. W., 1958: Introduction to Multivariate Statistical Analysis. John Wiley and Sons, Inc., New York, New York.

Anderson, T. W., 1963: The use of factor analysis in the statistical analysis of multiple time series. Psychometrika, 28, 1-25.

Barry, R. G., and A. H. Perry, 1973: Synoptic Climatology. Methuen and Co. LTD., 555 pp.

Brown, D. W., 1981: Tropical Storm Movement Forecasting Based on Synoptic Map Typing Using Empirical Orthogonal Functions. M. S. Thesis, Naval Postgraduate School, Monterey, California, 80 pp.

Cattell, R. B., 1958: Extracting the correct number of factors in factor analysis. Educational and Psychological Measurement, 18, 791-837.

Chatfield, C., 1980: The Analysis of Time Series: An Introduction, 2nd ed. Chapman and Hall, London and New York.

Craddock, J. M., and C. R. Flood, 1969: Eigenvectors for representing the 500mb geopotential surface over the Northern Hemisphere. Quarterly Journal of the Royal Meteorological Society, 95, 576-793.

Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomolies over the North Pacific Ocean. Journal of Physical Oceanography, 6, 249-266.

Dixon, W. J. and M. B. Brown, 1979: BMDP Biomedical Computer Programs P-Series, Univesrity of California Press, Berkeley, California, pp. 367-460.

Gerald, Curtis E., 1978: Applied Numerical Analysis, 2nd ed., Addison-Wesley Publishing Co., pp. 304-328 (Chapter 6).

Guttman, L., 1954: Some necessary conditions for common factor analysis. Psychometrika, 19, 149-161.

Haltiner, G. J. and R. T. Williams, 1980: Numerical Weather Prediction and Dynamic Meteorology. 2nd ed., John Wiley and Sons, Inc., New York, New York, 476 pp.

Harrison, E. J. Jr., 1973: Three-dimensional numerical simulation of tropical systems utilizing nested finite grids. Journal of Atmospheric Sciences, 30, 1528-1543.

--------. 1981: Initial results from the Navy two-way interactive nested tropical cyclone model. Monthly Weather Review, 109, 173-177.

Holton, J. D., 1972: An Introduction to Dynamic Meteorology. Academic Press, New York, 319 pp.

Horel, J. D., 1981: A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500mb height field. Monthly Weather Review, 103, 2080-2092.

Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441.

Huschke, R. E., 1959: Glossary of Meteorology. American Meteorological Society., Boston, Massachusetts.

Jarrell, J. D., and W. L. Sommervell Jr., 1970: A computer technique for using typhoon analogues as a forecast aid. NAVWEARSCHFAC Technical Paper Number 6-70, 47 pp.

Kidson, J. W., 1975: Eigenvector analysis of the monthly mean surface data. Monthly Weather Review, 103, 177-186.

Kutzbach, J. E., 1967: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. Journal of Applied Meteorology, 6, 791-802.

Leftwich, R. W. and C. J. Neumann, 1977: Statistical guidance for the prediction of eastern North Pacific tropical cyclone motion. Part 2. NOAA Technical Memo, WR-125, National Weather Service Western Region, Salt Lake City, Utah, 32 pp. (NTISPB-272 661/OGI).

Lorenz, E. N., 1956: Empircal Orthogonal Function and Statistical Weather Prediction. Massachusetts Institute of Technology Department of Meteorology, Cambridge, Massachusetts. Scientific Report 1, Statistical Forecasting Project, 48 pp.

---------, 1977: An experiment in non-linear statistical weather forecasting. Monthly Weather Review, 105, 590-602.

Morrison, D. F., 1967: Multivariate Statistical Methods. McGraw-Hill, New York, 338 pp. (ed. 2, 1976).

Neter, J., and W. Wasserman, 1974: Applied Linear Statistical Models. Richard D. Irwin, Inc., Homewood, Illinois, 842 pp.

Pearson, K., 1901: On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2, 559-572.

Peterson, K. A., 1980: Improvement in Tropical Cyclone Forecasts by Multiple Linear Regression Equation Adjustment of Analog Tracks. M. S. Thesis, Naval Postgraduate School. Monterey, California, 47 pp.

Preisendorfer, R. W., and T. P. Barnett, 1977: Significance tests for empirical orthogonal functions. Proceedings from the 5th conference on Probability and Statistics in Meteorology in Atmospheric Science, Nov. 15-18, 1977, American Meteorological Society, Boston, MA, 169-172.

Richman, M. B., 1980: Map Typing Patterns Associated with Urban Enhanced Precipitation. M. S. Thesis, University of Illinois, Urbana, 123 pp. (available from Illinois Department of Water Management, Champaign, Illinois).

---------, 1981: Obliquely rotated principal components: an improved map typing technique? Journal of Applied Meteorology, 20, 1145-1159.

Rinne, J., and V. Karhila, 1979: Empirical Orthogonal Functions of the 500mb height in the Northern Hemisphere determined from a large data sample. Quarterly Journal of the Royal Meteorological Society, 105, 873-884.

Stidd, C. K., 1967: The use of eigenvectors for climatic estimates. Journal of Applied Meteorology, 6, 255-264.

Walsh, J. E., and A. Mostek, 1980: A quantitative analysis of meteorological anomaly patterns over the United States, 1900-1977. Monthly Weather Review, 108, 615-630.

Walsh, J. E., and M. B. Richman, 1981: Seasonality in the association between surface temperature over the United States and the North Pacific Ocean. Monthly Weather Review, 109, 767-783.

INITIAL DISTRIBUTION LIST

No. Copies

1. Professor Russell L. Elsberry, Code 63Es 5
Department of Meteorology
Naval Postgraduate School
Monterey, California 93940

2. Capt. Alan R. Shaffer, USAF 5
Department of Meteorology, Code 63
Naval Postgraduate School
Monterey, California 93940

3. Defense Technical Information Center 2
Cameron Station
Alexandria, Virginia 22314

4. Library, Code 0142 2
Naval Postgraduate School
Monterey, California 93940

5. Air Weather Service 2
Technical Library
Scott Air Force Base, Illinois 62225

6. Dr. Robert J. Renard, Code 63Rd 1
Chairman, Department of Meteorology
Naval Postgraduate School
Monterey, California 93940

7. Captain Brian Van Orman 1
Program Manager, AFIT/CIRF
Air Force Institute of Technology
Wright-Patterson Air Force Base, Ohio 45433

8. Commanding Officer 1
Fleet Numerical Oceanography Center
Monterey, California 93940

9. Commanding Officer 1
Naval Environmental Prediction Research Facility
Monterey, California 93940

10. Commander 1
Air Weather Service
Scott Air Force Base,
St. Louis, Missouri 62225

11. Systems and Applied Sciences Corporation 1
570 Casanova Avenue
Monterey, California 93940

12. Lt. Scott Sandgathe, Code 63Ss 1
Department of Meteorology
Naval Postgraduate School
Monterey, California 93940

13. Lt. Dennis G. Larson 1
Geophysics Officer
USS Pelieu, LHA5
Long Beach, California 90801

14. Lt. Barry Donovan, USN 1
Oceanographic Centre
STAFF CINCEASTLANT
APO, New York 09241

Shaffer

Typhoon motion forecasting using empirical orthogonal function analysis of the synoptic forcing.

Thesis 197065
S4328 Shaffer
c.1
Typhoon motion forecasting using empirical orthogonal function analysis of the synoptic forcing.